

TWO-SAMPLE HYPOTHESIS TESTING FOR RANDOM DOT PRODUCT GRAPHS VIA ADJACENCY SPECTRAL EMBEDDING

BY MINH TANG, AVANTI ATHREYA, DANIEL L. SUSSMAN,
VINCE LYZINSKI AND CAREY E. PRIEBE*

Johns Hopkins University

We consider a semiparametric problem of two-sample hypothesis testing for a class of latent position random graphs. We formulate a notion of consistency in this context and propose a valid test for the hypothesis that two finite-dimensional random dot product graphs on a common vertex set have the same generating latent positions or have generating latent positions that are scaled or diagonal transformations of one another. Our test statistic is a function of a spectral decomposition of the adjacency matrix for each graph and our test procedure is consistent across a broad range of alternatives.

1. Introduction. Two-sample hypothesis testing for specific population parameters is a cornerstone of classical statistical inference. Furthermore, there exists a burgeoning literature on techniques for inference on random graphs, including vertex clustering, vertex classification and estimation of latent positions for a single graph [4, 7, 16, 17, 20, 21, 22]; and graph matching [8, 25, 28] and anomaly detection across multiple graphs [26]. We focus in this paper on a test for the hypothesis that two finite-dimensional random dot product graphs on the same vertex set, with known vertex correspondence, have the same generating latent position or have generating latent positions that are scaled or diagonal transformations of one another. We use the adjacency spectral embedding, which is a spectral decomposition of the adjacency matrix, for each graph to construct a computationally tractable, at most level- α test that is consistent across a broad collection of alternatives. Our proofs rely on estimates from [2, 16, 18, 22, 23, 24]. We note that we are able to obtain an at most level- α consistent test without specifying the finite-sample or asymptotic distribution of our test statistic. Gaussian distributional results of a related function of the adjacency spectral embedding are obtained in [2], under the more restrictive assumptions of independence of the rows of the latent position matrices. While we surmise that this distributional result will have implications for hypothesis testing, the situation we consider here is sufficiently distinct, and our methods in this paper do not require explicit knowledge of the asymptotic distribution of our test statistic. Test procedures similar to those described in this paper can also be constructed using other embedding methods, such as spectral decompositions of normalized Laplacian matrices. To prove the requisite bounds for these Laplacian-based test statistics, however, requires substantial technical machinery, and non-trivial adaptation or generalization of the results in [6, 19, 20], among others. Hence, for simplicity, we focus here on embeddings of the adjacency matrix.

*Supported in part by National Security Science and Engineering Faculty Fellowship (NSSEFF), Johns Hopkins University Human Language Technology Center of Excellence (JHU HLT COE), and the XDATA program of the Defense Advanced Research Projects Agency (DARPA) administered through Air Force Research Laboratory contract FA8750-12-2-0303.

AMS 2000 subject classifications: Primary 62G10; secondary 62H12, 05C80

Keywords and phrases: semiparametric graph inference, two-sample hypothesis testing, consistency, random dot product graph

In a slight generalization of our model, we can also view the latent positions as drawn from some pair of underlying distributions, say F and G , and hence the nonparametric test of equality of F and G is also of interest. We formulate this latter test, but determination of a suitable test statistic and its properties is the subject of ongoing research.

We emphasize that in the two-sample graph testing problem we address, the parameter dimension grows as the sample size grows. Hence this problem is not precisely analogous to classical two-sample tests for, say, the difference of two parameters belonging to some fixed Euclidean space, in which an increase in data has no effect on the dimension of the parameter. The problem is also not strictly nonparametric, since we view our latent positions as fixed parameters and thus impose specific distributional requirements on the data—that is, on the adjacency matrices. Indeed, we regard the problem as semiparametric, and we adapt the traditional definition of consistency to this setting. In particular, we have power increasing to one for alternatives in which the difference between the two latent positions grows as the sample size grows.

We illustrate our results through numerical simulations, and we remark on extensions and limitations of these results. We conclude the paper with a discussion of the applicability of other test statistics, including an intuitively appealing test based on the spectral norm of the difference of adjacency matrices. The specific hypotheses we consider are indicative of the multitude of questions that can arise in the larger context of two-sample hypothesis testing on random graphs.

2. Setting. We focus here on two-sample hypothesis testing for the latent position vectors of a pair of *random dot product graphs* (RDPG)[27] on the same vertex set with a known vertex correspondence, i.e., a bijective map φ from the vertex set of one graph to the vertex set of the other graph. We shall assume, without loss of generality, that φ is the identity map. Random dot product graphs are a specific example of *latent position random graphs* [12], in which each vertex is associated with a latent position and, conditioned on the latent positions, the presence or absence of all edges in the graph are independent. The edge presence probability is given by a *link* function, which is a symmetric function of the two latent positions. We observe that latent position graphs are identical to *exchangeable random graphs* [1, 14]; both are conditionally edge-independent. Statistical analysis for random graphs has received much recent interest: see [10, 11] for reviews of the pertinent literature.

The link function in a random dot product graph is simply the dot product: the probability of an edge between two vertices is given by the dot product of the latent positions. It was shown in [23] that spectral decompositions of the adjacency matrix for a random dot product graph provide accurate estimates of the underlying latent positions. In this work, we extend the results in [16, 23, 24] to test three different hypotheses on a pair of random dot product graphs on the same vertex set.

We begin with a number of necessary definitions and notational conventions. First, we define a random dot product graph on \mathbb{R}^d as follows.

DEFINITION 2.1 (Random Dot Product Graph (RDPG)). Let χ_d^n be defined by

$$\chi_d^n = \{\mathbf{U} \in \mathbb{R}^{n \times d} : \mathbf{U}\mathbf{U}^T \in [0, 1]^{n \times n} \text{ and } \text{rank}(\mathbf{U}) = d\}$$

and let $\mathbf{X} = [X_1, \dots, X_n]^T \in \chi_d^n$. Suppose \mathbf{A} is a random adjacency matrix given by

$$\mathbb{P}[\mathbf{A}|\mathbf{X}] = \prod_{i < j} (X_i^T X_j)^{\mathbf{A}_{ij}} (1 - X_i^T X_j)^{1 - \mathbf{A}_{ij}}$$

Then we say that \mathbf{A} is the adjacency matrix of a *random dot product graph* with *latent position* \mathbf{X} of rank d .

We define the matrix $\mathbf{P} = (p_{ij})$ of edge probabilities by $\mathbf{P} = \mathbf{X}\mathbf{X}^T$. We will also write $\mathbf{A} \sim \text{Bernoulli}(\mathbf{P})$ to represent that the existence of an edge between any two vertices i, j , where $i > j$, is a Bernoulli random variable with probability p_{ij} ; edges are independent, and \mathbf{A} is symmetric. We emphasize that the random graphs we consider are undirected and loop-free.

Suppose we are given two adjacency matrices \mathbf{A}_1 and \mathbf{A}_2 for a pair of independent random dot product graphs on the same vertex set. We assume throughout this paper that the graphs are independent. Our goal is to develop a consistent, at most level- α test to determine whether or not the two generating latent positions are equal, up to an orthogonal transformation. Indeed, if $\mathcal{O}(d)$ represents the collection of orthogonal matrices in $\mathbb{R}^{d \times d}$ and if $\mathbf{W} \in \mathcal{O}(d)$, then $\mathbf{X}\mathbf{W}\mathbf{W}^T\mathbf{X}^T = \mathbf{P}$, leading to obvious non-identifiability.

Formally, we state the following two-sample testing problems for random dot product graphs. Let $\mathbf{X}_n, \mathbf{Y}_n \in \chi_d^n$ and define $\mathbf{P}_n = \mathbf{X}_n\mathbf{X}_n^T$ and $\mathbf{Q}_n = \mathbf{Y}_n\mathbf{Y}_n^T$. As before, let $\mathcal{O}(d)$ denote the set of orthogonal $d \times d$ matrices with real entries. Given $\mathbf{A} \sim \text{Bernoulli}(\mathbf{P}_n)$ and $\mathbf{B} \sim \text{Bernoulli}(\mathbf{Q}_n)$, we consider the following tests:

(a) (*Equality, up to an orthogonal transformation*)

$$\begin{aligned} & H_0^n: \mathbf{X}_n \perp \mathbf{Y}_n \\ \text{against } & H_a^n: \mathbf{X}_n \not\perp \mathbf{Y}_n \end{aligned}$$

where \perp denotes that there exists an orthogonal matrix $\mathbf{W} \in \mathbb{R}^{d \times d}$ such that $\mathbf{X}_n = \mathbf{Y}_n\mathbf{W}$.

(b) (*Scaling*)

$$\begin{aligned} & H_0^n: \mathbf{X}_n \perp c_n \mathbf{Y}_n \text{ for some constant } c_n > 0 \\ \text{against } & H_a^n: \mathbf{X}_n \not\perp c_n \mathbf{Y}_n \text{ for any constant } c_n > 0 \end{aligned}$$

(c) (*Diagonal Transformation*)

$$\begin{aligned} & H_0^n: \mathbf{X}_n \perp \mathbf{D}_n \mathbf{Y}_n \text{ for some diagonal } \mathbf{D}_n \text{ with bounded positive entries} \\ \text{against } & H_a^n: \mathbf{X}_n \not\perp \mathbf{D}_n \mathbf{Y}_n \text{ for any diagonal } \mathbf{D}_n \text{ with bounded positive entries.} \end{aligned}$$

(d) (*Equality of underlying distributions*) Lastly, we state a nonparametric problem. Let the rows of $\mathbf{X}_n \in \chi_d^n$ be independent samples drawn from some common distribution F and the rows of $\mathbf{Y}_n \in \chi_d^n$ be independent samples drawn from some common distribution G . We wish to test

$$H_0: F = G \quad \text{against} \quad H_a: F \neq G$$

We stress that in our formulation of (a) – (c), the latent positions $\mathbf{X}_n, \mathbf{Y}_n$ need not be related to $\mathbf{X}_{n'}, \mathbf{Y}_{n'}$ for any $n' \neq n$. However, the size of the adjacency matrices \mathbf{A} and \mathbf{B} is quadratic in n and hence the larger n is, the more accurate are our estimates of \mathbf{X}_n and \mathbf{Y}_n . To contextualize our choice of hypotheses, consider the specific case of the stochastic blockmodel [13] and the related degree-corrected stochastic blockmodel [15]. In (a), we test whether two stochastic blockmodel graphs G_1 and G_2 with fixed block assignments have the same block probability matrices $\mathbf{N}_1 = \mathbf{N}_2$.

In (b), we test whether the block probability matrix of one graph is a scalar multiple of the other; i.e. if $\mathbf{N}_1 = c\mathbf{N}_2$. Finally, in (c), we test whether two degree-corrected stochastic blockmodel have the same block probability matrices (but possibly different degree-correction factors).

We now describe the adjacency spectral embedding. Given a latent position \mathbf{X} , the random dot product model generates an independent-edge adjacency matrix \mathbf{A} with edge probabilities given by $\mathbf{P} = \mathbf{X}\mathbf{X}^T$. We define the following spectral decomposition of \mathbf{P} by

$$\mathbf{P} = [\mathbf{U}_{\mathbf{P}}|\tilde{\mathbf{U}}_{\mathbf{P}}][\mathbf{S}_{\mathbf{P}} \oplus \tilde{\mathbf{S}}_{\mathbf{P}}][\mathbf{U}_{\mathbf{P}}|\tilde{\mathbf{U}}_{\mathbf{P}}],$$

where $\mathbf{S}_{\mathbf{P}} \in \mathbb{R}^{d \times d}$ is the diagonal matrix of the d largest eigenvalues of \mathbf{P} , and $[\mathbf{U}_{\mathbf{P}}|\tilde{\mathbf{U}}_{\mathbf{P}}] \in \mathbb{R}^{n \times n}$ is an orthonormal matrix in which $\mathbf{U}_{\mathbf{P}} \in \mathbb{R}^{n \times d}$ is the matrix whose columns are the d eigenvectors corresponding to the eigenvalues in $\mathbf{S}_{\mathbf{P}}$.

We will assume that the nonzero eigenvalues of S_P are distinct, ordered, and bounded from below, namely

$$\mathbf{S}_{\mathbf{P}}(1, 1) \geq \mathbf{S}_{\mathbf{P}}(2, 2) \geq \dots \geq \mathbf{S}_{\mathbf{P}}(d, d) > 0$$

and furthermore, since \mathbf{P} is rank d , we have that $\tilde{\mathbf{S}}_{\mathbf{P}} = 0$. Since P is unknown, however, this decomposition cannot be explicitly computed. Hence, we compute the analogous spectral decomposition for the adjacency matrix:

$$\mathbf{A} = [\mathbf{U}_{\mathbf{A}}|\tilde{\mathbf{U}}_{\mathbf{A}}][\mathbf{S}_{\mathbf{A}} \oplus \tilde{\mathbf{S}}_{\mathbf{A}}][\mathbf{U}_{\mathbf{A}}|\tilde{\mathbf{U}}_{\mathbf{A}}].$$

Observe that in the context of an RDPG, we can assume that the diagonal entries of $\mathbf{S}_{\mathbf{A}}$ are strictly positive. We use this decomposition of the adjacency matrix to define the *adjacency spectral embedding* of \mathbf{A} (see [22]):

DEFINITION 2.2. The *adjacency spectral embedding* of A is given by $\hat{\mathbf{X}} = \mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{1/2}$.

Let \mathbf{X} and \mathbf{Y} be two latent positions in $\mathbb{R}^{n \times d}$, and let $\mathbf{A} \sim \text{Bernoulli}(\mathbf{P})$ with $\mathbf{P} = \mathbf{X}\mathbf{X}^T$ and $\mathbf{B} \sim \text{Bernoulli}(\mathbf{Q})$ with $\mathbf{Q} = \mathbf{Y}\mathbf{Y}^T$ represent the associated adjacency matrices of the random dot product graphs with \mathbf{X} and \mathbf{Y} , respectively, as their latent positions. We observe that \mathbf{X} , \mathbf{Y} , and \mathbf{A} and \mathbf{B} all depend on n , but for notational convenience we will suppress this dependence except when imperative for communicating an asymptotic property. Let $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ denote the corresponding adjacency spectral embeddings of \mathbf{A} and \mathbf{B} , respectively. We use $\|\cdot\|_F$ to denote the Frobenius norm of a matrix; $\|\cdot\|$ to denote the spectral norm of a matrix or the Euclidean norm of a vector, depending on the context, and $\|\cdot\|_{2 \rightarrow \infty}$ to denote the $2 \rightarrow \infty$ operator norm (that is, the maximum Euclidean norm of the rows of a matrix). Also, we define, for a matrix \mathbf{M} of edge probabilities, the parameters $\delta(\mathbf{M})$, $\gamma_1(\mathbf{M})$, and $\gamma_2(\mathbf{M})$ as follows

$$\begin{aligned} \delta(\mathbf{M}) &= \max_{1 \leq i \leq n} \sum_{j=1}^n M_{ij} \\ \gamma_1(\mathbf{M}) &= \min_{i, j \in [d+1], i \neq j} \frac{|\mathbf{S}_{\mathbf{M}}(i, i) - \mathbf{S}_{\mathbf{M}}(j, j)|}{\delta(\mathbf{M})} \\ \gamma_2(\mathbf{M}) &= \min_{1 \leq i \leq d, d+1 \leq j \leq n} \frac{|\mathbf{S}_{\mathbf{M}}(i, i) - \mathbf{S}_{\mathbf{M}}(j, j)|}{\delta(\mathbf{M})} \end{aligned}$$

We remark that $\delta(\mathbf{P})$ is simply the maximum expected degree of a graph $\mathbf{A} \sim \text{Bernoulli}(\mathbf{P})$, $\gamma_1(\mathbf{P})$ is the minimum gap between distinct eigenvalues of \mathbf{P} , normalized by the maximum expected degree, and lastly, because \mathbf{P} is of rank d , $\gamma_2(\mathbf{P})$ is just $\mathbf{S}_P(d, d)/\delta(\mathbf{P})$. It is immediate that $\gamma_1 \leq \gamma_2$.

We rely on a number of established bounds on the separation between \mathbf{A} and \mathbf{P} and the accuracy of the adjacency spectral embedding in the estimation of the true latent positions. To be specific, we state below the requisite assumptions and several useful bounds which are proved in [2, 3, 18, 23, 24].

PROPOSITION 2.3. *Let $\hat{\mathbf{X}}_n \in \mathbb{R}^{n \times d}$ be the adjacency spectral embedding of the $n \times n$ adjacency matrix $\mathbf{A}_n \sim \text{Bernoulli}(\mathbf{P}_n)$ where $\mathbf{P}_n = \mathbf{X}_n \mathbf{X}_n^T$ is of rank d and its non-zero eigenvalues are distinct. Suppose also that there exists $\epsilon > 0$ such that $\delta(\mathbf{P}_n) \geq (\log n)^{1+\epsilon}$. Let $c > 0$ be arbitrary but fixed. There exists $n_0(c)$ such that if $n > n_0$ and η satisfies $n^{-c} < \eta < 1/2$, then with probability at least $1 - \eta$,*

$$(2.1) \quad \|\mathbf{P}_n - \mathbf{A}_n\| \leq 2\sqrt{\delta(\mathbf{P}_n) \log(n/\eta)}$$

and

$$(2.2) \quad \min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}_n \mathbf{W} - \mathbf{X}_n\|_F \leq 4\gamma_2^{-1}(\mathbf{P}_n) \sqrt{2d \log(n/\eta)}$$

Throughout this work, our results depend on certain growth conditions on the gap between the eigenvalues of \mathbf{P}_n and certain minimum sparsity conditions on \mathbf{P}_n as n increases. In light of Proposition 2.3, we consolidate our eigengap and sparsity assumptions on the sequence \mathbf{P}_n as follows:

ASSUMPTION 1. *We assume that there exists $d \in \mathbb{N}$ such that for all n , \mathbf{P}_n is of rank d . Further, we assume that there exist constants $\epsilon > 0$, $c_0 > 0$ and $n_0(\epsilon, c) \in \mathbb{N}$ such that for all $n \geq n_0$:*

$$(2.3) \quad \gamma_1(\mathbf{P}_n) > c_0$$

$$(2.4) \quad \delta(\mathbf{P}_n) > (\log n)^{2+\epsilon}$$

Because the parameters $\delta(\mathbf{P})$, $\gamma_1(\mathbf{P})$ and $\gamma_2(\mathbf{P})$ depend on \mathbf{P} , they cannot be computed from the adjacency matrices alone. Therefore, we use the corresponding estimates of these quantities, namely $\delta(\mathbf{A})$, $\gamma_1(\mathbf{A})$, and $\gamma_2(\mathbf{A})$. The following proposition, which is a straightforward consequence of Proposition 2.3, Hoeffding's inequality, and the Borel-Cantelli lemma, establishes that these estimates are, in fact, accurate.

PROPOSITION 2.4. *Let $\{\mathbf{X}_n\}$ be a sequence of latent positions and suppose that the sequence of matrices $\{\mathbf{P}_n\}$, where $\mathbf{P}_n = \mathbf{X}_n \mathbf{X}_n^T$, satisfy the conditions in Assumption 1. Let $\{\mathbf{A}_n\}$ be the sequence of adjacency matrices $\mathbf{A}_n \sim \text{Bernoulli}(\mathbf{P}_n)$. Then we have*

$$(2.5) \quad \frac{\delta(\mathbf{A}_n)}{\delta(\mathbf{P}_n)} \xrightarrow{\text{a.s.}} 1; \quad \frac{\gamma_1(\mathbf{A}_n)}{\gamma_1(\mathbf{P}_n)} \xrightarrow{\text{a.s.}} 1; \quad \frac{\gamma_2(\mathbf{A}_n)}{\gamma_2(\mathbf{P}_n)} \xrightarrow{\text{a.s.}} 1;$$

We note that a level- α test can easily be generated from Eq. (2.2) itself. In this paper we provide an improved concentration inequality, stated below, for $\hat{\mathbf{X}}$; in addition to having interesting distributional implications for $\hat{\mathbf{X}}$, the concentration inequality allows us to obtain better power over a larger class of alternatives.

THEOREM 2.5. *Suppose $\mathbf{P} = \mathbf{X}\mathbf{X}^T$ is an $n \times n$ probability matrix of rank d and its eigenvalues are distinct. Suppose also that there exists $\epsilon > 0$ such that $\delta(\mathbf{P}) > (\log n)^{2+\epsilon}$. Let $c > 0$ be arbitrary but*

fixed. Then there exists a $n_0(c)$ and a universal constant $C \geq 0$ such that if $n \geq n_0$ and $n^{-c} < \eta < 1/2$, then there exists a deterministic $\mathbf{W} \in \mathcal{O}(d)$ such that, with probability at least $1 - 3\eta$,

$$(2.6) \quad \left| \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}\|_F - C(\mathbf{X}) \right| \leq \frac{Cd^{3/2} \log(n/\eta)}{C(\mathbf{X})\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}}$$

where $C(\mathbf{X})$ is a constant depending only on \mathbf{X} and is bounded from above by $\sqrt{d\gamma_2^{-1}(\mathbf{P})}$.

REMARK. We observe that $C(\mathbf{X})$ is a function of the unknown latent position \mathbf{X} ; the exact expression is given in Lemma 4.4. We shall only require an upper bound on $C(\mathbf{X})$ for our subsequent results. Nevertheless, $C(\mathbf{X})$ can be consistently estimated, as $n \rightarrow \infty$, by a function $C(\hat{\mathbf{X}})$ of the $\hat{\mathbf{X}}$ as follows

$$C(\hat{\mathbf{X}}) = \sqrt{\text{tr } \mathbf{S}_A^{-1} \hat{\mathbf{X}}^T \hat{\mathbf{D}} \hat{\mathbf{X}} \mathbf{S}_A^{-1}}$$

where $\hat{\mathbf{D}}$ is the diagonal matrix whose diagonal entries are given by

$$\hat{\mathbf{D}}_{ii} = \sum_{j \neq i} \langle \hat{X}_i, \hat{X}_j \rangle (1 - \langle \hat{X}_i, \hat{X}_j \rangle)$$

Using $C(\hat{\mathbf{X}})$ in place of the upper bound for $C(\mathbf{X})$ in Theorem 2.5 may yield, in practice, better performance for the test procedures of Section 3.

As a corollary of Theorem 2.5, we obtain the following.

COROLLARY 2.6. *Let $\{\mathbf{X}_n\}$ be a sequence of latent positions and suppose that the sequence of matrices $\{\mathbf{P}_n\}$ where $\mathbf{P}_n = \mathbf{X}_n \mathbf{X}_n^T$ satisfies Assumption 1. Then there exists a deterministic sequence of orthogonal matrices \mathbf{W}_n such that*

$$\|\hat{\mathbf{X}}_n - \mathbf{X}_n \mathbf{W}_n\|_F - C(\mathbf{X}_n) \xrightarrow{\text{a.s.}} 0$$

We define our test statistic in the next section. However, we remark that based on current bounds, as n grows, the test statistic we construct will not always distinguish between two latent positions \mathbf{X}_n and \mathbf{Y}_n that differ in a constant number of rows. To this end, we introduce a slightly different notion of consistency in this context. For the test $\mathbf{X}_n \perp \mathbf{Y}_n$, this is given as follows.

DEFINITION 2.7. Let $\{\mathbf{X}_n\}, \{\mathbf{Y}_n\}$ be two sequences of latent positions $n \in \mathbb{N}$ satisfying Assumption 1 and such that

$$\min_{\mathbf{W} \in \mathcal{O}(d)} \|\mathbf{X}_n \mathbf{W} - \mathbf{Y}_n\| \rightarrow \infty$$

A test statistic T_n and associated rejection region R_n to test the null hypothesis

$$H_0^n : \mathbf{X}_n \perp \mathbf{Y}_n$$

against $H_a^n : \mathbf{X}_n \not\perp \mathbf{Y}_n$

is *consistent* if it rejects H_0^n with probability p_n , where $p_n \rightarrow 1$ as $n \rightarrow \infty$.

Thus, we have power converging to one over the class of alternatives in which the orthogonal Procrustes distance between the latent positions diverges.

3. Main Results. The first result is concerned with finite sample and asymptotic properties of a test for the null hypothesis $H_0 : \mathbf{X}_n \perp \mathbf{Y}_n$ against the alternative $H_a : \mathbf{X}_n \not\perp \mathbf{Y}_n$, for both the finite sample case of a fixed pair of latent positions \mathbf{X}_n and \mathbf{Y}_n and the asymptotic case of a sequence of latent positions $\{\mathbf{X}_n, \mathbf{Y}_n\}$, $n \in \mathbb{N}$. Our test statistic T_n is, in essence, a scaled version of

$$\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}_n \mathbf{W} - \hat{\mathbf{Y}}_n\|_F;$$

if it is sufficiently small, we do not reject; and if it is larger than a constant for which we provide an upper bound, then we reject.

THEOREM 3.1. *For each fixed n , consider the hypothesis test*

$$H_0^n : \mathbf{X}_n \perp \mathbf{Y}_n \quad \text{versus} \quad H_a^n : \mathbf{X}_n \not\perp \mathbf{Y}_n$$

where \mathbf{X}_n and $\mathbf{Y}_n \in \mathbb{R}^{n \times d}$ are matrices of latent positions for two random dot product graphs. Let $\hat{\mathbf{X}}_n$ and $\hat{\mathbf{Y}}_n$ be the adjacency spectral embeddings of $\mathbf{A}_n \sim \text{Bernoulli}(\mathbf{P}_n = \mathbf{X}_n \mathbf{X}_n^T)$ and $\mathbf{B}_n \sim \text{Bernoulli}(\mathbf{Q}_n = \mathbf{Y}_n \mathbf{Y}_n^T)$, respectively. Define the test statistic T_n as follows:

$$(3.1) \quad T_n = \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}_n \mathbf{W} - \hat{\mathbf{Y}}_n\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{A}_n)} + \sqrt{d\gamma_2^{-1}(\mathbf{B}_n)}}.$$

Let $\alpha \in (0, 1)$ be given. Then for all $C > 1$, if the rejection region is defined by

$$R := \{t \in \mathbb{R} : t \geq C\},$$

then there exists an $n_1 = n_1(\alpha, C) \in \mathbb{N}$ such that for all $n \geq n_1$, the test procedure with T_n and rejection region R is an at most level α test, i.e., for all $n \geq n_1$, if $\mathbf{X}_n \perp \mathbf{Y}_n$, then

$$\mathbb{P}(T_n \in R) \leq \alpha.$$

Furthermore, this sequence of tests satisfies the corresponding notion of consistency given in Definition 2.7.

PROOF. For ease of notation, in parts of the proof below we will suppress the dependence of \mathbf{X}_n , \mathbf{Y}_n , \mathbf{P}_n and \mathbf{Q}_n on n and simply denote these matrices by \mathbf{X} , \mathbf{Y} , \mathbf{P} , and \mathbf{Q} , respectively; we will make this dependence explicit when necessary. Suppose that the null hypothesis H_0 is true, so there exists an orthogonal $\widetilde{\mathbf{W}} \in \mathbb{R}^{d \times d}$ such that $\mathbf{X} = \mathbf{Y} \widetilde{\mathbf{W}}$. Let α be given, and let $\eta < \alpha/4$. From (2.6), for all n sufficiently large, there exist orthogonal matrices \mathbf{W}_X and $\mathbf{W}_Y \in \mathcal{O}(d)$ such that with probability at least $1 - \eta$,

$$\begin{aligned} \|\hat{\mathbf{X}} - \mathbf{X} \mathbf{W}_Y\|_F &\leq C(\mathbf{X}) + f(\mathbf{X}, \alpha, n) \\ \|\hat{\mathbf{Y}} - \mathbf{Y} \mathbf{W}_Y\|_F &\leq C(\mathbf{Y}) + f(\mathbf{Y}, \alpha, n) \end{aligned}$$

where $f(\mathbf{X}, \alpha, n) \rightarrow 0$ as $n \rightarrow \infty$ for a fixed \mathbf{X} and α .

Let $\mathbf{W}^* = \mathbf{W}_Y \widetilde{\mathbf{W}} \mathbf{W}_X$. Then there exists a $n_0 = n_0(\alpha)$ such that for all $n > n_0$, with probability at least $1 - \eta$, we have

$$\begin{aligned} \|\hat{\mathbf{X}} - \hat{\mathbf{Y}} \mathbf{W}^*\|_F &\leq \|\hat{\mathbf{X}} - \mathbf{X} \mathbf{W}_X\|_F + \|\mathbf{X} \mathbf{W}_X - \mathbf{Y} \widetilde{\mathbf{W}} \mathbf{W}_X\|_F + \|\mathbf{Y} \widetilde{\mathbf{W}} \mathbf{W}_X - \hat{\mathbf{Y}} \mathbf{W}^*\|_F \\ &\leq \|\hat{\mathbf{X}} - \mathbf{X} \mathbf{W}_X\|_F + \|\mathbf{X} \mathbf{W}_X - \mathbf{Y} \widetilde{\mathbf{W}} \mathbf{W}_X\|_F + \|(\mathbf{Y} - \hat{\mathbf{Y}} \mathbf{W}_Y)(\widetilde{\mathbf{W}} \mathbf{W}_X)\|_F \\ &\leq \|\hat{\mathbf{X}} - \mathbf{X} \mathbf{W}_X\|_F + \|\mathbf{Y} - \hat{\mathbf{Y}} \mathbf{W}_Y\|_F \\ &\leq C(\mathbf{X}) + C(\mathbf{Y}) + f(\mathbf{X}, \alpha, n) + f(\mathbf{Y}, \alpha, n) \end{aligned}$$

where we have used the fact that under H_0 , $\mathbf{X} = \mathbf{Y}\widetilde{\mathbf{W}}$; the final step follows from this and the unitary invariance of the Frobenius norm. We note that both $C(\mathbf{X})$ and $C(\mathbf{Y})$ are unknown. However, by Theorem 2.5, they can be bounded from above by $(d\gamma_2^{-1}(\mathbf{P}))^{1/2}$ and $(d\gamma_2^{-1}(\mathbf{Q}))^{1/2}$, respectively. Hence for all $n > n_0$, with probability at least $1 - \alpha$, we have

$$\frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}_n \mathbf{W} - \hat{\mathbf{Y}}_n\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{P}_n)} + \sqrt{d\gamma_2^{-1}(\mathbf{Q}_n)}} \leq 1 + r(\alpha, n)$$

where $r(\alpha, n) \rightarrow 0$ as $n \rightarrow \infty$ for a fixed α . In addition, by Proposition 2.4, the terms $\gamma_2^{-1}(\mathbf{P}_n)$ and $\gamma_2^{-1}(\mathbf{Q}_n)$ in the denominator can be replaced by $\gamma_2^{-1}(\mathbf{A}_n)$ and $\gamma_2^{-1}(\mathbf{B}_n)$ for sufficiently large n . Therefore, with probability at least $1 - \alpha$,

$$T_n = \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}_n \mathbf{W} - \hat{\mathbf{Y}}_n\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{A}_n)} + \sqrt{d\gamma_2^{-1}(\mathbf{B}_n)}} \leq 1 + \tilde{r}(\alpha, n)$$

where once again, for a fixed α , $\tilde{r}(\alpha, n) \rightarrow 0$ as $n \rightarrow \infty$. We can thus take $n_1 = n_1(\alpha, C) = \inf\{n \geq n_0(\alpha) : \tilde{r}(\alpha, n) \leq C - 1\} < \infty$. Then for all $n > n_1$ and $\mathbf{X}_n, \mathbf{Y}_n$ satisfying $\mathbf{X}_n \perp \mathbf{Y}_n$, we conclude

$$\mathbb{P}(T_n \in R) < \alpha.$$

We now prove consistency. Let $\widetilde{\mathbf{W}} = \operatorname{argmin}_{\mathbf{W} \in \mathcal{O}(d)} \|\mathbf{X} - \mathbf{Y}\mathbf{W}\|_F$ and denote by $D(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\widetilde{\mathbf{W}}\|_F$. As before, let $\mathbf{W}^* = \mathbf{W}_Y \widetilde{\mathbf{W}} \mathbf{W}_X$. Note that

$$\begin{aligned} (3.2) \quad \|\hat{\mathbf{X}} - \hat{\mathbf{Y}}\mathbf{W}^*\|_F &\geq \|\mathbf{Y}\widetilde{\mathbf{W}} - \mathbf{X}\|_F - \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F + \|\mathbf{Y}\mathbf{W}_Y\mathbf{W} - \hat{\mathbf{Y}}\mathbf{W}\|_F \\ &\geq D(\mathbf{X}, \mathbf{Y}) - \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F - \|\mathbf{Y}\mathbf{W}_Y - \hat{\mathbf{Y}}\|_F \end{aligned}$$

Therefore, for all n ,

$$\begin{aligned} \mathbb{P}(T_n \notin R) &\leq \mathbb{P}\left(\frac{\|\hat{\mathbf{X}} - \hat{\mathbf{Y}}\mathbf{W}^*\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{A})} + \sqrt{d\gamma_2^{-1}(\mathbf{B})}} \leq C\right) \\ &\leq \mathbb{P}\left(\frac{\|\mathbf{Y}\widetilde{\mathbf{W}} - \mathbf{X}\|_F - \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F - \|\mathbf{Y}\mathbf{W}_Y - \hat{\mathbf{Y}}\|_F}{\sqrt{d\gamma_2^{-1}(\mathbf{A})} + \sqrt{d\gamma_2^{-1}(\mathbf{B})}} \leq C\right) \\ &= \mathbb{P}\left(\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F + \|\mathbf{Y}\mathbf{W}_Y - \hat{\mathbf{Y}}\|_F + C' \geq D(\mathbf{X}, \mathbf{Y})\right) \end{aligned}$$

where $C' = C(\sqrt{d\gamma_2^{-1}(\mathbf{A})} + \sqrt{d\gamma_2^{-1}(\mathbf{B})})$. By Assumption 1, there exists some n_0 and some $c_0 > 0$ such that $\gamma_2(\mathbf{P}_n) \geq c_0$ and $\gamma_2(\mathbf{Q}_n) \geq c_0$ for all $n \geq n_0$. Now, let $\beta > 0$ be given. By the almost sure convergence of $\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_x\|_F$ to $C(\mathbf{X})$, established in Theorem 2.5, and the almost sure convergence of $\gamma_2(\mathbf{A})$ to $\gamma_2(\mathbf{P})$ given in 2.4, we deduce that there exists a constant $M_1(\beta)$ and a positive integer $n_0 = n_0(\alpha, \beta)$ so that, for all $n \geq n_0(\alpha, \beta)$,

$$\begin{aligned} \mathbb{P}(\|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}_X\|_F + C\sqrt{d\gamma_2^{-1}(\mathbf{A})} \geq M_1/2) &\leq \beta/2 \\ \mathbb{P}(\|\hat{\mathbf{Y}} - \mathbf{Y}\mathbf{W}_Y\|_F + C\sqrt{d\gamma_2^{-1}(\mathbf{B})} \geq M_1/2) &\leq \beta/2 \end{aligned}$$

Since $D(\mathbf{X}_n, \mathbf{Y}_n) \rightarrow \infty$ under the alternative, there exists some $n_2 = n_2(\alpha, \beta, C)$ such that, for all $n \geq n_2$, $D(\mathbf{X}_n, \mathbf{Y}_n) \geq M_1$. Hence, for all $n \geq n_2$, $\mathbb{P}(T_n \notin R) \leq \beta$, i.e., our test statistic T_n lies within the rejection region R with probability at least $1 - \beta$, as required. \square

Let $\mathcal{C} = \mathcal{C}(\mathbf{Y}_n)$ denote the class of all positive constants c for which all the entries of $c^2 \mathbf{Y}_n \mathbf{Y}_n^T$ belong to the unit interval. We next consider the case of testing the null hypothesis $H_0: \mathbf{X}_n \perp c_n \mathbf{Y}_n$ for some $c_n \in \mathcal{C}$ against the alternative that $H_a: \mathbf{X}_n \not\perp c_n \mathbf{Y}_n$ for any $c_n \in \mathcal{C}$. In what follows below, we will only write $c_n > 0$, but will always assume that $c_n \in \mathcal{C}$, since the problem is ill-posed otherwise. The test statistic T_n is now a simple modification of the one used in Theorem 3.1: for this test, we compute a Procrustes distance between scaled adjacency spectral embeddings for the two graphs. Specifically, we have

THEOREM 3.2. *For each fixed n , consider the hypothesis test*

$$\begin{aligned} H_0^n: \mathbf{X}_n &\perp c_n \mathbf{Y}_n \quad \text{for some } c_n > 0 \text{ versus} \\ H_a^n: \mathbf{X}_n &\not\perp c_n \mathbf{Y}_n \quad \text{for all } c_n > 0 \end{aligned}$$

where \mathbf{X}_n and $\mathbf{Y}_n \in \mathbb{R}^{n \times d}$ are latent positions for two random dot product graphs with adjacency matrices \mathbf{A}_n and \mathbf{B}_n , respectively. Define the test statistic T_n as follows:

$$(3.3) \quad T_n = \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}_n \mathbf{W} / \|\hat{\mathbf{X}}_n\|_F - \hat{\mathbf{Y}}_n / \|\hat{\mathbf{Y}}_n\|_F\|_F}{2\sqrt{d\gamma_2^{-1}(\mathbf{A}_n)} / \|\hat{\mathbf{X}}_n\|_F + 2\sqrt{d\gamma_2^{-1}(\mathbf{B}_n)} / \|\hat{\mathbf{Y}}_n\|_F}.$$

Let $\alpha \in (0, 1)$ be given. Then for all $C > 1$, if the rejection region is defined by

$$R := \{t \in \mathbb{R} : t \geq C\},$$

then there exists an $n_1 = n_1(\alpha, C) \in \mathbb{N}$ such that for all $n \geq n_1$, the test procedure with T_n and rejection region R is an at most level α test. Furthermore, if $\{\mathbf{X}_n\}$ and $\{\mathbf{Y}_n\}$, $n \in \mathbb{N}$, are two sequences of rank d latent positions each of which satisfies Assumption 1 and, in addition, if

$$(3.4) \quad \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\mathbf{X}_n \mathbf{W} / \|\mathbf{X}_n\|_F - \mathbf{Y}_n / \|\mathbf{Y}_n\|_F\|_F}{C(\mathbf{X}_n) / \|\mathbf{X}_n\|_F + C(\mathbf{Y}_n) / \|\mathbf{Y}_n\|_F} \rightarrow \infty$$

then the sequence of tests satisfies a corresponding notion of consistency; namely, for any $\beta > 0$ there exists a $n_2(\alpha, \beta, C)$ such that $\inf_{n \geq n_2} \mathbb{P}(T_n \in R) \geq 1 - \beta$.

REMARK. We remark that the collection of alternatives in Eq. (3.4) is, effectively, those latent positions \mathbf{X}_n and \mathbf{Y}_n whose scalings onto the sphere remain far enough apart as $n \rightarrow \infty$. Indeed, we note that the denominator of our test statistic converges to zero, so we require that the numerator does not become small too quickly. Moreover, since $C(\mathbf{X}_n)$ and $C(\mathbf{Y}_n)$ are bounded by from above by constants, we can replace them by fixed constants (say, 1) in the denominator of Eq. (3.4) and obtain an equivalent class of alternatives.

PROOF. The proof of this result is almost identical to that of Theorem 3.1. We sketch here the necessary modifications. As before, we suppress dependence on n unless necessary. Let α be given and let $\eta = \alpha/4$. By Theorem 2.5, for n sufficiently large, there exists some orthogonal $\mathbf{W}_{\mathbf{X}} \in \mathcal{O}(d)$ such that, with probability at least $1 - \eta$

$$\|\hat{\mathbf{X}} \mathbf{W}_{\mathbf{X}} - \mathbf{X}\|_F \leq C(\mathbf{X}) + f(\mathbf{X}, \alpha, n)$$

where for any fixed \mathbf{X} and α , $f(\mathbf{X}, \alpha, n) \rightarrow 0$ as $n \rightarrow \infty$. From the form of $C(\mathbf{X})$ given in Theorem 2.5, for all n ,

$$(3.5) \quad C(\mathbf{X}_n) \leq \sqrt{\frac{d}{\gamma_2(\mathbf{P}_n)}}$$

Now, again for n sufficiently large,

$$\begin{aligned}
\|\hat{\mathbf{X}}\mathbf{W}_{\mathbf{X}}/\|\hat{\mathbf{X}}\|_F - \mathbf{X}/\|\mathbf{X}\|_F\|_F &\leq \frac{\|\hat{\mathbf{X}}\mathbf{W}_{\mathbf{X}} - \mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} + \|\mathbf{X}\|_F \left| \frac{1}{\|\hat{\mathbf{X}}\|_F} - \frac{1}{\|\mathbf{X}\|_F} \right| \\
&\leq \frac{\|\hat{\mathbf{X}}\mathbf{W}_{\mathbf{X}} - \mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} + \frac{\|\hat{\mathbf{X}}\|_F - \|\mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} \\
&\leq \frac{\|\hat{\mathbf{X}}\mathbf{W}_{\mathbf{X}} - \mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} + \frac{\|\hat{\mathbf{X}}\mathbf{W}_{\mathbf{X}}\|_F - \|\mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} \\
&\leq \frac{2\|\hat{\mathbf{X}}\mathbf{W}_{\mathbf{X}} - \mathbf{X}\|_F}{\|\hat{\mathbf{X}}\|_F} \leq \frac{2(C(\mathbf{X}) + f(\mathbf{X}, \alpha, n))}{\|\hat{\mathbf{X}}\|_F}
\end{aligned}$$

with probability at least $1 - \eta$. An analogous bound can also be derived for \mathbf{Y} . Under the null hypothesis, $\mathbf{X} \perp c\mathbf{Y}$ for some $c > 0$, so we derive that

$$\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}\mathbf{W}/\|\hat{\mathbf{X}}\|_F - \hat{\mathbf{Y}}/\|\hat{\mathbf{Y}}\|_F\|_F \leq \frac{2(C(\mathbf{X}) + f(\mathbf{X}, \alpha, n))}{\|\hat{\mathbf{X}}\|_F} + \frac{2(C(\mathbf{Y}) + f(\mathbf{Y}, \alpha, n))}{\|\hat{\mathbf{Y}}\|_F}$$

From Eq.(3.5) and Proposition 2.4, we conclude that for n sufficiently large,

$$T_n = \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}\mathbf{W}/\|\hat{\mathbf{X}}\|_F - \hat{\mathbf{Y}}/\|\hat{\mathbf{Y}}\|_F\|_F}{2\sqrt{d\gamma_2^{-1}(\mathbf{A})}/\|\hat{\mathbf{X}}\|_F + 2\sqrt{d\gamma_2^{-1}(\mathbf{B})}/\|\hat{\mathbf{Y}}\|_F} \leq 1 + r(\alpha, n)$$

where $r(\alpha, n) \rightarrow 0$ as $n \rightarrow \infty$ for a fixed α . We can now choose a $n_1 = n_1(\alpha, C)$ for which $r(\alpha, n_1) \leq C - 1$. This implies that for all $n \geq n_1$,

$$\mathbb{P}(T_n \in R) \leq \alpha$$

which establishes that the test statistic T_n with rejection region R is an at most level- α test. The proof of consistency proceeds in an almost identical manner to that in Theorem 3.1 and we omit the details. \square

Finally, we consider the case of testing the null hypothesis $H_0: \mathbf{X}_n \perp \mathbf{D}_n \mathbf{Y}_n$ for some diagonal matrix \mathbf{D}_n . We proceed analogously to the previous case by defining the class $\mathcal{E} = \mathcal{E}(\mathbf{Y}_n)$ to be all positive diagonal matrices $\mathbf{D}_n \in \mathbb{R}^{n \times n}$ such that $\mathbf{D}_n \mathbf{Y}_n \mathbf{Y}_n^T \mathbf{D}_n$ has all entries in the unit interval. Once more, we focus on testing the null hypothesis $H_0: \mathbf{X}_n \perp \mathbf{D}_n \mathbf{Y}_n$ for some $\mathbf{D}_n \in \mathcal{E}$ against the alternative $H_a: \mathbf{X}_n \not\perp \mathbf{D}_n \mathbf{Y}_n$ for any matrix $\mathbf{D}_n \in \mathcal{E}$. As before, in what follows in this section, we will always assume that \mathbf{D}_n belongs to \mathcal{E} , even if this assumption is not explicitly stated. The test statistic T_n in this case is again a simple modification of the one used in Theorem 3.1. However, for technical reasons, our proof of consistency requires an additional condition on the minimum Euclidean norm of each row of the matrices \mathbf{X}_n and \mathbf{Y}_n . To avoid certain technical issues, we impose a slightly stronger density assumption on our graphs for this test. These assumptions can be weakened, but at cost of interpretability. The assumptions we make on the latent positions, which we summarize here, are moderate restrictions the sparsity of the graphs.

ASSUMPTION 2. *We assume that there exists $d \in \mathbb{N}$ such that for all n , \mathbf{P}_n is of rank d . Further, we assume that there exist constants $\epsilon_1 > 0$, $\epsilon_2 > 0$, $c_0 > 0$ and $n_0(\epsilon_1, \epsilon_2, c) \in \mathbb{N}$ such that for all*

$n \geq n_0$:

$$(3.6) \quad \gamma_1(\mathbf{P}_n) > c_0$$

$$(3.7) \quad \delta(\mathbf{P}_n) > n^{1/2}(\log n)^{\epsilon_1}$$

$$(3.8) \quad \min_i \|X_i\| > \left(\frac{\log n}{\sqrt{\delta(\mathbf{P}_n)}} \right)^{1-\epsilon_2}$$

We then have the following result.

THEOREM 3.3. *For each fixed n , consider the hypothesis test*

$$\begin{aligned} H_0^n: \mathbf{X}_n \perp \mathbf{D}_n \mathbf{Y}_n & \text{ for some diagonal } \mathbf{D}_n \in \mathcal{E} \text{ versus} \\ H_a^n: \mathbf{X}_n \not\perp \mathbf{D}_n \mathbf{Y}_n & \text{ for any diagonal } \mathbf{D}_n \in \mathcal{E} \end{aligned}$$

where \mathbf{X}_n and $\mathbf{Y}_n \in \mathbb{R}^{n \times d}$ are matrices of latent positions for two random dot product graphs. Let $\hat{\mathbf{X}}_n$ and $\hat{\mathbf{Y}}_n$ be the adjacency spectral embeddings for \mathbf{X} and \mathbf{Y} , respectively. For any matrix $\mathbf{Z} \in \mathbb{R}^{n \times d}$, let $\mathcal{D}(\mathbf{Z})$ be the diagonal matrix whose diagonal entries are the Euclidean norm of the rows of \mathbf{Z} , i.e.,

$$\mathcal{D}(\mathbf{Z}) = (\text{diag}(\mathbf{Z}\mathbf{Z}^T))^{1/2}.$$

Denote by $\mathcal{P}(\mathbf{Z})$ the projection of the rows of \mathbf{Z} onto the unit sphere, i.e.,

$$\mathcal{P}(\mathbf{Z}) = \mathcal{D}^{-1}(\mathbf{Z})\mathbf{Z}$$

where, for simplicity of notation, we write $\mathcal{D}^{-1}(\mathbf{Z})$ for $(\mathcal{D}(\mathbf{Z}))^{-1}$. For any matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$ with positive entries, define $S(\mathbf{M})$ as follows:

$$(3.9) \quad S(\mathbf{M}) = \frac{170d^{3/2} \log(n/\eta)}{\sqrt{\gamma_1^7(\mathbf{M})\delta(\mathbf{M})}}.$$

We define the test statistic as follows:

$$(3.10) \quad T_n = \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\mathcal{P}(\hat{\mathbf{X}}_n)\mathbf{W} - \mathcal{P}(\hat{\mathbf{Y}}_n)\|_F}{S(\mathbf{A}_n)\|\mathcal{D}^{-1}(\hat{\mathbf{X}}_n)\|_F + S(\mathbf{B}_n)\|\mathcal{D}^{-1}(\hat{\mathbf{Y}}_n)\|_F}.$$

Let $\alpha \in (0, 1)$ be given. Then for all $C > 1$, if the rejection region is defined by

$$R := \{t \in \mathbb{R} : t \geq C\},$$

then there exists an $n_1 = n_1(\alpha, C) \in \mathbb{N}$ such that for all $n \geq n_1$, the test procedure with T_n and rejection region R is an at most level- α test. Furthermore, if $\{\mathbf{X}_n\}$ and $\{\mathbf{Y}_n\}$, $n \in \mathbb{N}$, are two sequences of rank d latent positions each of which satisfies Assumption 2 and, in addition, if

$$(3.11) \quad D_{\mathcal{P}}(\mathbf{X}, \mathbf{Y}) := \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\mathcal{P}(\mathbf{X}_n)\mathbf{W} - \mathcal{P}(\mathbf{Y}_n)\|_F}{S(\mathbf{P}_n)\|\mathcal{D}^{-1}(\mathbf{X})\|_F + S(\mathbf{Q}_n)\|\mathcal{D}^{-1}(\mathbf{Y})\|_F} \rightarrow \infty$$

then the sequence of tests also satisfies a corresponding notion of consistency; namely, for any $\beta > 0$ there exists a $n_2(\alpha, \beta, C)$ such that $\inf_{n \geq n_2} \mathbb{P}(T_n \in R) \geq 1 - \beta$.

REMARK. If the latent positions of \mathbf{X} and \mathbf{Y} are related by a diagonal transformation, this implies that each row X_i of \mathbf{X} is a scaled version of the corresponding row Y_i of \mathbf{Y} ; that is, $X_i = c_i Y_i$. Under the null, the angle between the adjacency spectral embeddings \hat{X}_i and \hat{Y}_i should be small. This suggests that we consider a cosine distance between the rows, and the projection in the numerator of our test statistic is essentially just that: namely, it measures the distance between projections of rows of the latent positions on the sphere (see Fig. 1). We make this specific choice of test statistic because it allows us to exploit a series of existing results on bounds for the deviations between true and estimated latent positions. There are several other reasonable choices of test statistic; our happens to be straightforward to analyze, and the denominator is a natural consequence of a collection of known bounds on the accuracy of the adjacency spectral embedding. Furthermore, this explains our choice for the class of alternatives, which may initially appear rather Byzantine, for which the test satisfies a notion of consistency similar to that given in Definition 2.7. To illustrate, consider the following example of the class of alternatives for which Eq. (3.11) is satisfied. Let \mathbf{X}_n and \mathbf{Y}_n be two matrices of latent positions and suppose that $\min_i \|X_i\| \geq c$ and $\min_i \|Y_i\| \geq c$ for some constant $c > 0$. Then $\delta(\mathbf{P}_n)$ and $\delta(\mathbf{Q}_n)$ are of order $\Theta(n)$. $S(\mathbf{P}_n)$ and $S(\mathbf{Q}_n)$ are of order $\Theta(\log n / \sqrt{n})$. Finally, $\|\mathcal{D}^{-1}(\mathbf{X}_n)\|_F$ and $\|\mathcal{D}^{-1}(\mathbf{Y}_n)\|_F$ are of order $\Theta(\sqrt{n})$. Therefore, Eq. (3.11) is satisfied provided that $\min_{\mathbf{W}} \|\mathcal{P}(\mathbf{X}_n)\mathbf{W} - \mathcal{P}(\mathbf{Y}_n)\|_F$ diverges to ∞ at a rate faster than $\log n$. We also emphasize that $D_{\mathcal{P}}(\mathbf{X}, \mathbf{Y})$ has the property $D_{\mathcal{P}}(c\mathbf{X}, c\mathbf{Y}) = c^2 D_{\mathcal{P}}(\mathbf{X}, \mathbf{Y})$ for all $c \in (0, 1]$. This suggests that our test procedure may not have much power when the graphs are sparse.

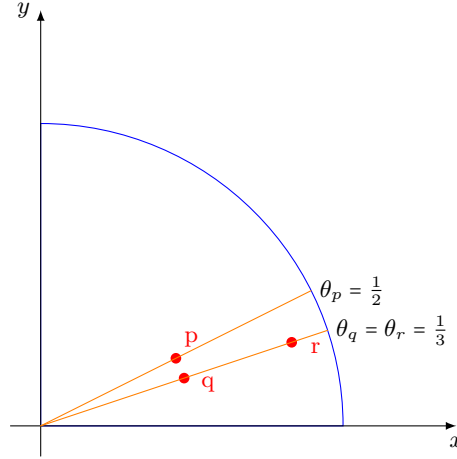


Fig 1: A pictorial example to illustrate the effect of projection. The distance between p and q is originally small, but increases after projection of p to $\theta_p = 1/2$ and q to $\theta_q = 2/5$. The distance between q and r after projection is zero and the distance between p and r after projection decreases.

Our proof of Theorem 3.3 relies on a variant of Theorem 2.5, namely the following bound on the maximum of the l_2 norm of the rows of $\hat{\mathbf{X}} - \mathbf{X}$.

LEMMA 3.4. *Suppose Assumption 2 holds, and let $c > 0$ be arbitrary. Then there exists a $n_0(c)$ such that for all $n > n_0$ and $n^{-c} < \eta < 1/2$, there exists a deterministic $\mathbf{W} = \mathbf{W}_n \in \mathcal{O}(d)$ such that, with probability at least $1 - 3\eta$,*

$$(3.12) \quad \|\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}\|_{2 \rightarrow \infty} = \max_i \|\hat{X}_i - \mathbf{W}X_i\| \leq \frac{85d^{3/2} \log(n/\eta)}{\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}}.$$

PROOF OF THEOREM 3.3. For any $\mathbf{W} \in \mathcal{O}(d)$,

$$\begin{aligned} \|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\mathbf{X})\|_F &= \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\hat{\mathbf{X}}\mathbf{W} - \mathcal{D}^{-1}(\hat{\mathbf{X}})\mathbf{X} + \mathcal{D}^{-1}(\hat{\mathbf{X}})\mathbf{X} - \mathcal{D}^{-1}(\mathbf{X})\mathbf{X}\|_F \\ &\leq \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F \|\hat{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2 \rightarrow \infty} + \|(\mathcal{D}^{-1}(\hat{\mathbf{X}}) - \mathcal{D}^{-1}(\mathbf{X}))\mathbf{X}\|_F \end{aligned}$$

The term $\|(\mathcal{D}^{-1}(\hat{\mathbf{X}}) - \mathcal{D}^{-1}(\mathbf{X}))\mathbf{X}\|_F$ can now be written as

$$\begin{aligned} \|(\mathcal{D}^{-1}(\hat{\mathbf{X}}) - \mathcal{D}^{-1}(\mathbf{X}))\mathbf{X}\|_F^2 &= \sum_{i=1}^n \|X_i\|^2 \left(\frac{1}{\|\hat{X}_i\|} - \frac{1}{\|X_i\|} \right)^2 \\ &= \sum_{i=1}^n \frac{(\|\mathbf{W}X_i\| - \|\hat{X}_i\|)^2}{\|\hat{X}_i\|^2} \\ &\leq \sum_{i=1}^n \frac{\|\mathbf{W}X_i - \hat{X}_i\|^2}{\|\hat{X}_i\|^2} \\ &\leq (\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F \|\hat{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2 \rightarrow \infty})^2 \end{aligned}$$

and hence,

$$\|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\mathbf{X})\|_F \leq 2\|\hat{\mathbf{X}}\mathbf{W} - \mathbf{X}\|_{2 \rightarrow \infty} \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F$$

Now, let $\alpha \in (0, 1)$ be given and set $\eta = \alpha/4$. Then by Lemma 3.4, for n sufficiently large, there exists some orthogonal $\mathbf{W}_\mathbf{X}$ such that, with probability at least $1 - \eta$,

$$(3.13) \quad \|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W}_\mathbf{X} - \mathcal{P}(\mathbf{X})\|_F \leq \frac{170d^{3/2} \log(n/\eta)}{\sqrt{\gamma_1^7(\mathbf{P})}\delta(\mathbf{P})} \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F = S(\mathbf{P}) \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F.$$

An analogous bound holds for \mathbf{Y} . Therefore, under the null hypothesis $\mathbf{X} \perp \mathbf{D}\mathbf{Y}$, we have

$$(3.14) \quad \min_{\mathbf{W} \in \mathcal{O}(d)} \|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\hat{\mathbf{Y}})\|_F \leq S(\mathbf{P}) \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{Q}) \|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F$$

and hence,

$$\frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\hat{\mathbf{Y}})\|_F}{S(\mathbf{P}) \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{Q}) \|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F} \leq 1$$

Now, by Proposition 2.4, for sufficiently large n , we can replace $S(\mathbf{P})$ and $S(\mathbf{Q})$ with $S(\mathbf{A})$ and $S(\mathbf{B})$ to yield

$$T_n = \frac{\min_{\mathbf{W} \in \mathcal{O}(d)} \|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W} - \mathcal{P}(\hat{\mathbf{Y}})\|_F}{S(\mathbf{A}) \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{B}) \|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F} \leq 1 + r(\alpha, n)$$

where $r(\alpha, n) \rightarrow 0$ as $n \rightarrow \infty$ for a fixed α . We can therefore choose a $n_1 = n_1(\alpha, C)$ for which $r(\alpha, n_1) \leq C - 1$. This implies that for all $n \geq n_1$,

$$\mathbb{P}(T_n \in C) \leq \alpha$$

yielding that the test statistic T_n with rejection region R is an at most level- α test.

We now prove consistency of this test procedure. Suppose now that the sequence of latent positions $\{\mathbf{X}_n, \mathbf{Y}_n\}$ is such that $\mathbf{X}_n \not\perp \mathbf{D}_n \mathbf{Y}_n$ for any diagonal $\mathbf{D}_n \in \mathcal{E}$. Denote by $h(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$ the ratio

$$h(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) = \frac{S(\mathbf{P}) \|\mathcal{D}^{-1}(\mathbf{X})\|_F + S(\mathbf{Q}) \|\mathcal{D}^{-1}(\mathbf{Y})\|_F}{S(\mathbf{A}) \|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{B}) \|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F}$$

and by $f(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$ the ratio

$$f(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) = \frac{\|\mathcal{P}(\hat{\mathbf{X}})\mathbf{W}_{\mathbf{X}} - \mathcal{P}(\mathbf{X})\|_F + \|\mathcal{P}(\hat{\mathbf{Y}})\mathbf{W}_{\mathbf{Y}} - \mathcal{P}(\mathbf{Y})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{B})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F}$$

Then, for all n ,

$$\begin{aligned} \mathbb{P}(T_n \notin R_n) &\leq \mathbb{P}\left(\frac{\min_{W \in \mathcal{O}(d)} \|\mathcal{P}(\mathbf{X})\mathbf{W} - \mathcal{P}(\mathbf{Y})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F + S(\mathbf{B})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F} \leq C + f(\hat{\mathbf{X}}, \hat{\mathbf{Y}})\right) \\ &\leq \mathbb{P}\left(h(\hat{\mathbf{X}}, \hat{\mathbf{Y}})(C + f(\hat{\mathbf{X}}, \hat{\mathbf{Y}})) \geq D_{\mathcal{P}}(\mathbf{X}, \mathbf{Y})\right) \end{aligned}$$

Now, for a given $\beta > 0$, let $M_1 = M_1(\beta)$ and $n_0 = n_0(\alpha, \beta)$ be such that, for all $n \geq n_0(\alpha, \beta)$,

$$\mathbb{P}(C + f(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) \geq M_1) \leq \beta/2.$$

By Eq. (3.13) and Proposition 2.4, $M_1(\beta)$ and $n_0(\alpha, \beta)$ exists for all choice of β .

We now show that there exists, for any $\beta > 0$, some $n_1 = n_1(\beta)$ such that, for all $n \geq n_1(\beta)$,

$$(3.15) \quad \mathbb{P}(h(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) \geq 4) \leq \beta/2.$$

Indeed,

$$h(\hat{\mathbf{X}}, \hat{\mathbf{Y}}) \leq \max\left\{\frac{S(\mathbf{P})\|\mathcal{D}^{-1}(\mathbf{X})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F}, \frac{S(\mathbf{Q})\|\mathcal{D}^{-1}(\mathbf{Y})\|_F}{S(\mathbf{B})\|\mathcal{D}^{-1}(\hat{\mathbf{Y}})\|_F}\right\}$$

In addition, we have

$$\frac{\|\mathcal{D}^{-1}(\mathbf{X})\|_F}{\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F} = \frac{\sqrt{\sum_{i=1}^n 1/\|X_i\|^2}}{\sqrt{\sum_{i=1}^n 1/\|\hat{X}_i\|^2}} \leq \sqrt{\max_i \frac{\|\hat{X}_i\|^2}{\|X_i\|^2}} = \max_i \frac{\|\hat{X}_i\|}{\|X_i\|} \leq 1 + \frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_{2 \rightarrow \infty}}{\min_i \|X_i\|}$$

By Lemma 3.4 and the condition in Assumption 2 on $\min_i \|X_i\|$, there exist some $n_1(\beta)$ such that for all $n \geq n_1(\beta)$,

$$\mathbb{P}\left(\frac{\|\hat{\mathbf{X}} - \mathbf{X}\|_{2 \rightarrow \infty}}{\min_i \|X_i\|} \geq 1\right) \leq \beta/8.$$

We now apply Proposition 2.4 to $S(\mathbf{P})/S(\mathbf{A})$ to conclude that there exist some $n_1(\beta)$ such that for all $n \geq n_1(\beta)$,

$$\mathbb{P}\left(\frac{S(\mathbf{P})\|\mathcal{D}^{-1}(\mathbf{X})\|_F}{S(\mathbf{A})\|\mathcal{D}^{-1}(\hat{\mathbf{X}})\|_F} \geq 4\right) \leq \beta/4.$$

The same argument can be applied to the ratio depending on $\hat{\mathbf{Y}}$ and \mathbf{Y} . This establishes that there exists some $n_1(\beta)$ such that Eq. (3.15) holds for all $n \geq n_1(\beta)$. Since $D_{\mathcal{P}}(\mathbf{X}_n, \mathbf{Y}_n) \rightarrow \infty$, there exists some $n_2 = n_2(\alpha, \beta, C)$ such that for all $n \geq n_2$, $D_{\mathcal{P}}(\mathbf{X}_n, \mathbf{Y}_n) \geq 4M_1(\beta)$. Hence for all $n \geq n_2$, $\mathbb{P}(T_n \notin R) \leq \beta$, i.e., the test statistic lies within the rejection region with probability at least $1 - \beta$, as required. \square

4. Additional Proofs. In this section, we present the proof of Theorem 2.5 and Lemma 3.4. Theorem 2.5 is, in essence, an improvement of Eq. (2.2) in Proposition 2.3 and is due to results that provide more accurate control on the difference between the eigenvalues and eigenvectors of \mathbf{A} and those of \mathbf{P} , i.e., they do not follow directly from Eq. (2.1) and applications of Weyl's inequality or the Davis-Kahan $\sin \Theta$ theorem [9]. More specifically, we state the following bounds on the difference between the spectrum of \mathbf{A} and that of \mathbf{P} as proved in [2].

LEMMA 4.1. *If the events in Proposition 2.3 occur, then*

$$(4.1) \quad \|\mathbf{S}_{\mathbf{A}} - \mathbf{S}_{\mathbf{P}}\| \leq \frac{18d \log(n/\eta)}{\gamma_1^2(\mathbf{P})}$$

$$(4.2) \quad \|\mathbf{U}_{\mathbf{A}}^T(\mathbf{U}_{\mathbf{A}} - \mathbf{U}_{\mathbf{P}})\|_F \leq \frac{10d \log(n/\eta)}{\gamma_1^2(\mathbf{P})\delta(\mathbf{P})}$$

Let \mathbf{W} be such that $\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{1/2} = \mathbf{X}\mathbf{W}$. We note that such a matrix \mathbf{W} always exists as $\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}\mathbf{U}_{\mathbf{P}}^T = \mathbf{P} = \mathbf{X}\mathbf{X}^T$. The proofs of both Theorem 2.5 and Lemma 3.4 proceed by bounding, in a series of technical lemmas, each of the terms in parentheses in the following decomposition of $\hat{\mathbf{X}} - \mathbf{X}\mathbf{W}$:

$$\begin{aligned} \hat{\mathbf{X}} - \mathbf{X}\mathbf{W} &= \mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{1/2} - \mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{1/2} \\ &= \mathbf{A}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{P}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2} \\ &= (\mathbf{A}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{A}}^{-1/2}) + (\mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}) + (\mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2} - \mathbf{P}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}) \end{aligned}$$

We state the following lemmas, whose proofs can be found in [16]. The first lemma bounds $\|(\mathbf{A}\mathbf{U}_{\mathbf{A}}\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{A}}^{-1/2})\|_F$ by viewing it as the difference after one step of the power method for \mathbf{A} when starting at $\mathbf{U}_{\mathbf{P}}$. The second lemma bounds $\|\mathbf{A}\mathbf{U}_{\mathbf{P}}(\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{S}_{\mathbf{P}}^{-1/2})\|_F$.

LEMMA 4.2. *If the events in Proposition 2.3 occur, then*

$$(4.3) \quad \|\hat{\mathbf{X}} - \mathbf{A}\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{A}}^{-1/2}\|_F \leq \frac{24\sqrt{2}d \log(n/\eta)}{\sqrt{\gamma_1^5(\mathbf{P})\delta(\mathbf{P})}}$$

LEMMA 4.3. *If the events in Proposition 2.3 occur, then*

$$(4.4) \quad \|\mathbf{A}\mathbf{U}_{\mathbf{P}}(\mathbf{S}_{\mathbf{A}}^{-1/2} - \mathbf{S}_{\mathbf{P}}^{-1/2})\|_F \leq \frac{18d^{3/2} \log(n/\eta)}{\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}}.$$

The proof of Lemma 3.4 proceeds by using Hoeffding's inequality to bound the $2 \rightarrow \infty$ norm of $(\mathbf{A} - \mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}$. To prove Theorem 2.5, an analogous bound for the Frobenius norm is required. Our last technical lemma provides a concentration bound for $\|(\mathbf{A} - \mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F$. We note that the proof of this lemma also relies on classical concentration inequalities; the argument we give is straightforward, if tedious.

LEMMA 4.4. *Let $\eta > 0$ be arbitrary. Then with probability at least $1 - 2\eta$, the events in Proposition 2.3 occur and furthermore,*

$$(4.5) \quad \left| \|(\mathbf{A} - \mathbf{P})\mathbf{U}_{\mathbf{P}}\mathbf{S}_{\mathbf{P}}^{-1/2}\|_F^2 - C^2(\mathbf{X}) \right| \leq \frac{14\sqrt{2}d \log(n/\eta)}{\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}}.$$

where $C^2(\mathbf{X})$ is a constant that depends only on \mathbf{X} . In addition, $C^2(\mathbf{X})$ can be written as

$$C^2(\mathbf{X}) = \text{tr } \mathbf{S}_\mathbf{P}^{-1/2} \mathbf{U}_\mathbf{P}^T \mathbb{E}[(\mathbf{A} - \mathbf{P})^2] \mathbf{U}_\mathbf{P} \mathbf{S}_\mathbf{P}^{-1/2} = \text{tr } \mathbf{S}_\mathbf{P}^{-1/2} \mathbf{U}_\mathbf{P}^T \mathbf{D} \mathbf{U}_\mathbf{P} \mathbf{S}_\mathbf{P}^{-1/2} \leq d\gamma_2^{-1}(\mathbf{P})$$

where \mathbf{D} is a diagonal matrix whose diagonal entries are given by

$$\mathbf{D}_{ii} = \sum_{k \neq i} \mathbf{P}_{ik}(1 - \mathbf{P}_{ik}).$$

PROOF. Let $Z = \|(\mathbf{A} - \mathbf{P})\mathbf{U}_\mathbf{P}\mathbf{S}_\mathbf{P}^{-1/2}\|_F^2$. Since our graphs are undirected and loop free, Z is a function of the $\binom{n}{2}$ independent random variables $\{\mathbf{A}_{ij}\}_{i < j}$. Let \mathbf{A} and \mathbf{A}' be two arbitrary adjacency matrices. Denote by $\mathbf{A}^{(kl)}$ the adjacency matrix obtained by replacing the (k, l) and (l, k) entries of \mathbf{A} by those of \mathbf{A}' . Let $Z_{kl} = \|(\mathbf{A}^{(kl)} - \mathbf{P})\mathbf{U}_\mathbf{P}\mathbf{S}_\mathbf{P}^{-1/2}\|_F^2$. The argument we employ is based on the following logarithmic Sobolev concentration inequality for $Z - \mathbb{E}[Z]$ [5, §6.4]. Namely,

THEOREM 4.5. Assume that there exists a constant $v > 0$ such that, with probability at least $1 - \eta$,

$$\sum_{k < l} (Z - Z_{kl})^2 \leq v.$$

Then for all $t > 0$,

$$\mathbb{P}[|Z - \mathbb{E}[Z]| > t] \leq 2e^{-t^2/(2v)} + \eta.$$

Let $\mathbf{V} = \mathbf{U}_\mathbf{P}\mathbf{S}_\mathbf{P}^{-1/2}$. For notational convenience, we denote the i -th row of \mathbf{V} by V_i . We shall also denote the inner product between vectors in Euclidean space by $\langle \cdot, \cdot \rangle$. The i -th row of the product $(\mathbf{A} - \mathbf{P})\mathbf{V}$ is simply a linear combination of the rows of \mathbf{V} , i.e.,

$$((\mathbf{A} - \mathbf{P})\mathbf{V})_i = \sum_{j=1}^n (\mathbf{A} - \mathbf{P})_{ij} V_j.$$

Hence,

$$Z = \|(\mathbf{A} - \mathbf{P})\mathbf{V}\|_F^2 = \sum_{i=1}^n \|((\mathbf{A} - \mathbf{P})\mathbf{V})_i\|^2 = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (\mathbf{A} - \mathbf{P})_{ij} (\mathbf{A} - \mathbf{P})_{ik} \langle V_j, V_k \rangle$$

As \mathbf{A} and $\mathbf{A}^{(kl)}$ differs possibly only in the (k, l) and (l, k) entries and that the entries of \mathbf{A} and \mathbf{A}' are binary variables, we have that if $(Z - Z_{kl})$ is non-zero, then

$$\begin{aligned} Z - Z_{kl} &= 2 \left(\sum_{j \neq l} (\mathbf{A} - \mathbf{P})_{kj} \langle V_j, V_l \rangle \right) + 2 \left(\sum_{j \neq k} (\mathbf{A} - \mathbf{P})_{lj} \langle V_j, V_k \rangle \right) + (1 - 2\mathbf{P}_{kl}) \langle V_l, V_k \rangle \\ &= 2 \sum_{j=1}^n (\mathbf{A} - \mathbf{P})_{kj} \langle V_j, V_l \rangle + 2 \sum_{j=1}^n (\mathbf{A} - \mathbf{P})_{lj} \langle V_j, V_k \rangle + c_{kl} \end{aligned}$$

where $c_{kl} = 2(\mathbf{A} - \mathbf{P})_{kl} \langle V_l, V_l \rangle + 2(\mathbf{A} - \mathbf{P})_{lk} \langle V_k, V_k \rangle + (1 - 2\mathbf{P}_{kl}) \langle V_l, V_k \rangle$. We then have

$$(Z - Z_{kl})^2 \leq 3(C_{kl}^{(1)} + C_{kl}^{(2)} + c_{kl}^2)$$

where $C_{kl}^{(1)}$ and $C_{kl}^{(2)}$ are given by

$$\begin{aligned} C_{kl}^{(1)} &= 4 \sum_{j_1=1}^n \sum_{j_2=1}^n (\mathbf{A} - \mathbf{P})_{kj_1} (\mathbf{A} - \mathbf{P})_{kj_2} \langle V_{j_1}, V_{j_1} \rangle \langle V_{j_2}, V_{j_2} \rangle = 4 \left[((\mathbf{A} - \mathbf{P})\mathbf{V}\mathbf{V}^T)_{kl} \right]^2 \\ C_{kl}^{(2)} &= 4 \sum_{j_1=1}^n \sum_{j_2=1}^n (\mathbf{A} - \mathbf{P})_{lj_1} (\mathbf{A} - \mathbf{P})_{lj_2} \langle V_{j_1}, V_{j_1} \rangle \langle V_{j_2}, V_{j_2} \rangle = 4 \left[((\mathbf{A} - \mathbf{P})\mathbf{V}\mathbf{V}^T)_{lk} \right]^2 \end{aligned}$$

As $C_{kl}^{(1)} = C_{lk}^{(2)}$, $c_{kl} = c_{lk}$, and $C_{kk}^{(1)} > 0$ for all l, k , we thus have

$$\sum_{k < l} (Z - Z_{kl})^2 \leq 3 \sum_{k < l} (C_{kl}^{(1)} + C_{kl}^{(2)} + c_{kl}^2) \leq 3 \sum_{k=1}^n \sum_{l=1}^n C_{kl}^{(1)} + \frac{3}{2} \sum_{k=1}^n \sum_{l=1}^n c_{kl}^2$$

We now consider each of the term in the above right hand side.

$$\begin{aligned} \sum_{k=1}^n \sum_{l=1}^n C_{kl}^{(1)} &= 4 \sum_{k=1}^n \sum_{l=1}^n \left[((\mathbf{A} - \mathbf{P})\mathbf{V}\mathbf{V}^T)_{kl} \right]^2 = 4 \|(\mathbf{A} - \mathbf{P})\mathbf{V}\mathbf{V}^T\|_F^2 \\ \sum_{k=1}^n \sum_{l=1}^n c_{kl}^2 &\leq 3 \sum_{k=1}^n \sum_{l=1}^n 4(\mathbf{A} - \mathbf{P})_{kl}^2 (\langle V_l, V_l \rangle^2 + \langle V_k, V_k \rangle^2) + 3 \sum_{k=1}^n \sum_{l=1}^n \langle V_l, V_k \rangle^2 \\ &= 6 \sum_{k=1}^n 4((\mathbf{A} - \mathbf{P})^2)_{kk} \langle V_k, V_k \rangle^2 + 3 \sum_{k=1}^n \sum_{l=1}^n \langle V_l, V_k \rangle^2 \\ &= 24 \sum_{k=1}^n ((\mathbf{A} - \mathbf{P})^2)_{kk} \langle V_k, V_k \rangle^2 + 3 \sum_{k=1}^n (\mathbf{V}\mathbf{V}^T \mathbf{V}\mathbf{V}^T)_{kk} \\ &= 24 \sum_{k=1}^n ((\mathbf{A} - \mathbf{P})^2)_{kk} \langle V_k, V_k \rangle^2 + 3 \|\mathbf{V}\mathbf{V}^T\|_F^2 \\ &\leq 24 \|(\mathbf{A} - \mathbf{P})^2\| \sum_{k=1}^n \langle V_k, V_k \rangle^2 + 3 \|\mathbf{V}\mathbf{V}^T\|_F^2 \\ &\leq 24 \|\mathbf{A} - \mathbf{P}\|^2 \|\text{diag}(\mathbf{V}\mathbf{V}^T)\|_F^2 + 3 \|\mathbf{V}\mathbf{V}^T\|_F^2 \end{aligned}$$

where the penultimate inequality of the above display follows from the fact that the diagonal elements of $(\mathbf{A} - \mathbf{P})^2$ is majorized by its eigenvalues. We therefore have

$$\begin{aligned} \sum_{k < l} (Z - Z_{kl})^2 &\leq (48 \|\mathbf{A} - \mathbf{P}\|^2 + \frac{9}{2}) \|\mathbf{V}\mathbf{V}^T\|_F^2 \\ &\leq 49 \|\mathbf{A} - \mathbf{P}\|^2 \|\mathbf{V}\mathbf{V}^T\|_F^2 \\ &= 49 \|\mathbf{A} - \mathbf{P}\|^2 \|\mathbf{S}_\mathbf{P}^{-1}\|_F^2 \\ &\leq 49 \|\mathbf{A} - \mathbf{P}\|^2 \frac{d}{(\gamma_2(\mathbf{P})\delta(\mathbf{P}))^2} \end{aligned}$$

By Proposition 2.3, for any $\eta > 0$, with probability at least $1 - \eta$,

$$\|\mathbf{A} - \mathbf{P}\|^2 \leq 4\delta(\mathbf{P}) \log(n/\eta)$$

Hence, for all $\eta > 0$, with probability at least $1 - \eta$,

$$(4.6) \quad \sum_{k < l} (Z - Z_{kl})^2 \leq \frac{196d \log(n/\eta)}{\gamma_2^2(\mathbf{P})\delta(\mathbf{P})}.$$

Denote by $v(\eta)$ the right hand side of the above display. We then have, by Theorem 4.5, that for all $t > 0$,

$$(4.7) \quad \mathbb{P}[|Z - \mathbb{E}[Z]| > t] \leq 2e^{-t^2/(2v(\eta))} + \eta$$

Setting t to be

$$t = \frac{14\sqrt{2d} \log(n/\eta)}{\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}}$$

yields $2e^{-t^2/(2v(\eta))} \leq \eta$ as desired.

Finally, we provide a bound for $\mathbb{E}[Z]$ in terms of the parameters $\gamma_2(\mathbf{P})$ and $\delta(\mathbf{P})$. We have

$$\mathbb{E}[Z] = \mathbb{E}[\|(\mathbf{A} - \mathbf{P})\mathbf{V}\|_F^2] = \mathbb{E}[\text{tr}(\mathbf{V}^T(\mathbf{A} - \mathbf{P})^2\mathbf{V})] = \text{tr}(\mathbf{V}^T\mathbb{E}[(\mathbf{A} - \mathbf{P})^2]\mathbf{V})$$

We note that

$$\mathbb{E}[(\mathbf{A} - \mathbf{P})^2]_{ij} = \mathbb{E}\left[\sum_k (\mathbf{A} - \mathbf{P})_{ik}(\mathbf{A} - \mathbf{P})_{kj}\right] = \begin{cases} 0 & \text{if } i \neq j \\ \sum_{k \neq i} \mathbf{P}_{ik}(1 - \mathbf{P})_{ik} & \text{if } i = j \end{cases}$$

Hence, $\delta(\mathbf{P})\mathbf{I} - \mathbb{E}[(\mathbf{A} - \mathbf{P})^2]$ is positive semidefinite. We thus have

$$\mathbb{E}[\|(\mathbf{A} - \mathbf{P})\mathbf{V}\|_F^2] \leq \delta \text{tr} \mathbf{V}^T \mathbf{V} \leq d\gamma_2^{-1}(\mathbf{P}).$$

which establishes the upper bound $C^2(\mathbf{X}) \leq d\gamma_2^{-1}(\mathbf{P})$ as required. \square

PROOF OF THEOREM 2.5. From Lemma 4.4, we have

$$\left| \|(\mathbf{A} - \mathbf{P})\mathbf{U}_\mathbf{P}\mathbf{S}_\mathbf{P}^{-1/2}\|_F^2 - C^2(\mathbf{X}) \right| \leq \frac{14\sqrt{2d}\log(n/\eta)}{\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}}$$

with probability at least $1 - 2\eta$. Now, $a \leq b + c$ implies $\sqrt{a} \leq \sqrt{b} + \frac{c}{2\sqrt{b}}$ and $a \geq b - c \geq 0$ implies $\sqrt{a} \geq \sqrt{b} - \frac{c}{\sqrt{b}}$. Hence

$$-\frac{14\sqrt{2d}\log(n/\eta)}{C(\mathbf{X})\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}} \leq \|(\mathbf{A} - \mathbf{P})\mathbf{U}_\mathbf{P}\mathbf{S}_\mathbf{P}^{-1/2}\|_F - C(\mathbf{X}) \leq \frac{7\sqrt{2d}\log(n/\eta)}{C(\mathbf{X})\gamma_2(\mathbf{P})\sqrt{\delta(\mathbf{P})}}$$

with probability at least $1 - 2\eta$. Applying Lemma 4.2 and Lemma 4.3 yield

$$-\frac{C_1 d^{3/2} \log(n/\eta)}{C(\mathbf{X})\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}} \leq \|\hat{\mathbf{X}} - \mathbf{P}\mathbf{U}_\mathbf{P}\mathbf{S}_\mathbf{P}^{-1/2}\|_F - C(\mathbf{X}) \leq \frac{C_2 d^{3/2} \log(n/\eta)}{C(\mathbf{X})\sqrt{\gamma_1^7(\mathbf{P})\delta(\mathbf{P})}}$$

for some constants $C_1, C_2 > 0$. As $\mathbf{P}\mathbf{U}_\mathbf{P}\mathbf{S}_\mathbf{P}^{-1/2} = \mathbf{X}\mathbf{W}$ for some $\mathbf{W} \in \mathcal{O}(d)$, Theorem 2.5 follows. \square

5. Simulations. In this section, we illustrate the hypothesis tests and test statistics of Section 2 through several simulated data examples. We set $\alpha = 0.05$ as the upper bound on the type 1 error throughout. We first consider the problem of testing the null hypothesis $H_0: \mathbf{X}_n \perp \mathbf{Y}_n$ against the alternative hypothesis $\mathbf{H}_A: \mathbf{X}_n \not\perp \mathbf{Y}_n$. We consider random graphs generated according to two stochastic blockmodels [13] with the same block membership probability vector $\boldsymbol{\pi}$ but different block probability matrices \mathbf{B}_0 and \mathbf{B}_1 , where

$$(5.1) \quad \mathbf{B}_0 = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}; \quad \mathbf{B}_1 = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{bmatrix}; \quad \boldsymbol{\pi} = (0.4, 0.6).$$

We evaluate the performance of the test statistics through Monte Carlo simulation. We consider a sequence of $n \in \{100, 200, \dots, 2000\}$. For each choice of n , we generate 1000 Monte Carlo replicates; in each replicated we sample three graphs. Two of the graphs are generated with block probability matrix \mathbf{B}_0 and the remaining graph is generated with block probability matrix \mathbf{B}_1 . To keep the

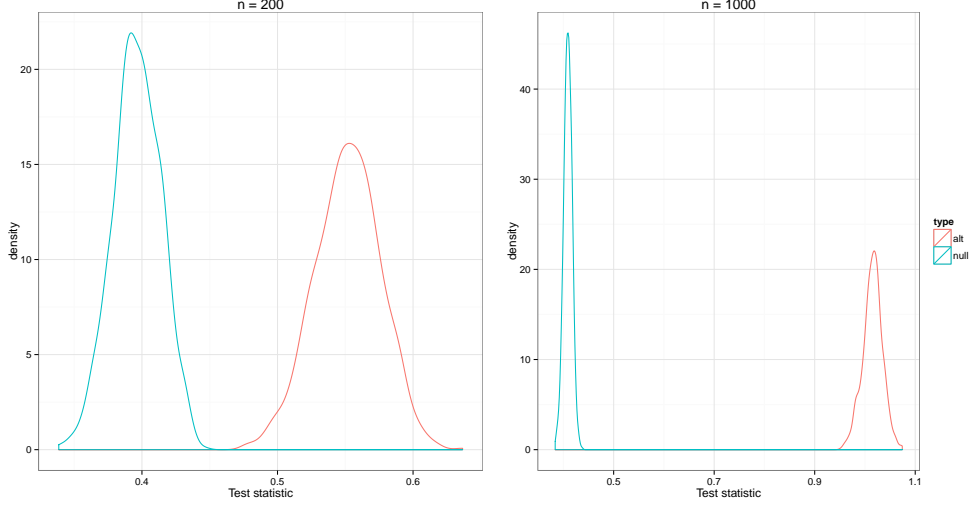


Fig 2: Density estimate for the test statistic when testing $H_0: \mathbf{X}_n \perp \mathbf{Y}_n$ against $H_A: \mathbf{X}_n \not\perp \mathbf{Y}_n$.

vertex set fixed and consistent across all three graphs, the block membership vector is sampled once for each Monte Carlo replicate. We use the graphs generated from \mathbf{B}_0 to evaluate the test statistics under the null hypothesis and we use a graph generated from \mathbf{B}_0 along with a graph generated from \mathbf{B}_1 for the alternative hypothesis. The Monte Carlo replicates for two values of n , namely $n = 200$ and $n = 1000$ are presented in Figure 2. We see from Figure 2 that the rejection region $R = \{T_n > 1\}$ as specified by Theorem 3.1 is very conservative. Indeed, the estimated power of the test using the theoretical rejection region when $n = 200$ is 0. However, the 95% percentile of T_n under the null is approximately 0.42; with this value as the critical value for the rejection region, the estimated power is 1. In addition, for $n = 100$, when the critical value is computed based on Monte Carlo replicates for the null, the estimated power is 0.864. The rejection region predicted by our theory becomes more relevant when $n \geq 1000$. For example, we obtain an estimated power of 0.80 Monte Carlo replicates and $n = 1000$ and rejection region as specified in Theorem 3.1, and corresponding estimated power of 1 when $n = 1200$.

We next consider the hypothesis test $H_0: \mathbf{X}_n \perp c_n \mathbf{Y}_n$ for some $c_n > 0$ against the alternative $H_A: \mathbf{X}_n \not\perp c_n \mathbf{Y}_n$ for any $c_n > 0$. We employ the model specified in Eq. (5.1). The distribution of the test statistic as given by the Monte Carlo replicates are presented in Figure 3 for two values of $n = 200$ and $n = 1000$. We see that the empirical power of the test is estimated to be roughly 0.15 for $n = 200$, which is much less powerful than that of the test in Figure 2, even though the random graphs models are identical. This is consistent with the notion that the null hypothesis considered in Figure 2 is a single element of the hypothesis space considered in Figure 3. For this setup, the theoretical rejection region as specified in Theorem 3.2 yields good power only for $n \geq 6000$.

Finally, we consider the hypothesis test $H_0: \mathbf{X}_n \perp \mathbf{D}_n \mathbf{Y}_n$ for some diagonal matrix \mathbf{D}_n with positive diagonal entries against the alternative $\mathbf{X}_n \not\perp \mathbf{D}_n \mathbf{Y}_n$ for all diagonal matrices \mathbf{D}_n with positive diagonal entries; in particular, we focus here on degree-corrected stochastic blockmodels [15] with block probability vector $\boldsymbol{\pi} = (0.4, 0.6)$ and block probability matrices \mathbf{B}_0 , \mathbf{B}_1 and \mathbf{B}_2 where

$$\mathbf{B}_0 = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}; \quad \mathbf{B}_2 = \begin{bmatrix} 0.72 & 0.192 \\ 0.192 & 0.32 \end{bmatrix} = \begin{bmatrix} 1.2 & 0 \\ 0 & 0.8 \end{bmatrix} \mathbf{B}_0 \begin{bmatrix} 1.2 & 0 \\ 0 & 0.8 \end{bmatrix}; \quad \mathbf{B}_1 = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{bmatrix}.$$

Recall that a degree corrected stochastic blockmodel graph G on n vertices with K blocks is

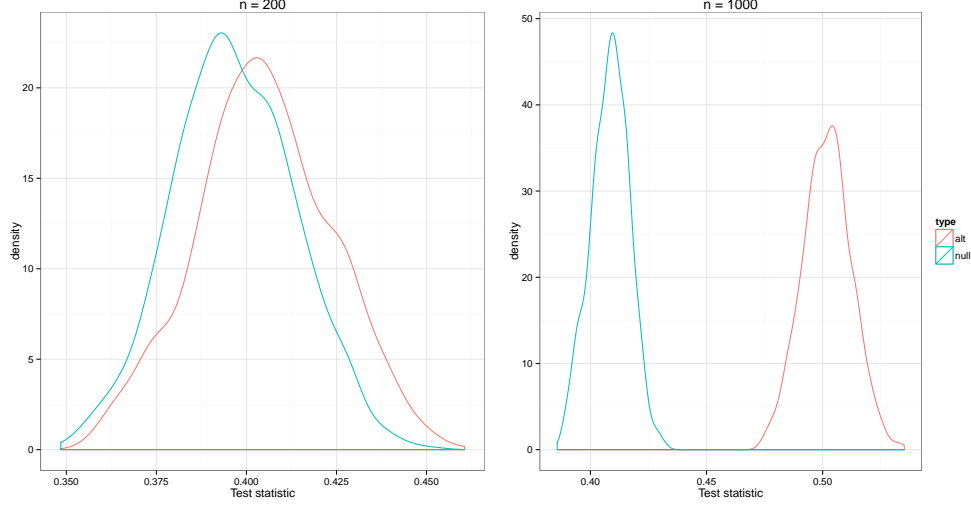


Fig 3: Density estimate for the test statistic when testing $H_0: \mathbf{X}_n \perp c_n \mathbf{Y}_n$ for some $c_n > 0$ against the alternative $H_A: \mathbf{X}_n \not\perp c_n \mathbf{Y}_n$ for all $c_n > 0$.

parametrized by a block probability vector $\pi \in \mathbb{R}^K$, a $K \times K$ block probability matrix \mathbf{B} , and a degree correction vector $\mathbf{c} \in \mathbb{R}^n$. The vertices of G are assigned into one of the K blocks. The edges of G are independent; furthermore, given that vertices i and j are assigned into block $\tau(i)$ and $\tau(j)$, the probability of an edge between i and j is simply $c_i c_j \mathbf{B}_{\tau(i), \tau(j)}$. The vector \mathbf{c} allows for heterogeneity of degree within blocks, in contrast to the homogeneity exhibited by traditional stochastic blockmodels.

By the above construction, \mathbf{B}_2 and \mathbf{B}_0 correspond to the same degree corrected stochastic blockmodel. In this case, we evaluate the performance of the test statistic by sampling, for each Monte Carlo replicate, three degree corrected stochastic block model graphs with underlying block probability matrices \mathbf{B}_0 , \mathbf{B}_1 and \mathbf{B}_2 . The degree correction for the vertices are i.i.d. draws from the uniform distribution on the interval $[0.2, 1]$. The graphs generated from \mathbf{B}_0 and \mathbf{B}_2 are used to compute the empirical distribution of the test statistic under the null, and the graphs generated from \mathbf{B}_0 and \mathbf{B}_1 are used to compute the empirical distribution under the alternative. The results are presented in Figure 4 for $n = 200$ and $n = 4000$. The test once again exhibits good power when using the rejection region obtained via the Monte Carlo replicates. As before, however, the theoretical rejection region is extremely conservative; for this choice of null and alternative hypothesis, the theoretical rejection region yields power only when n is at least of the order of 10^6 .

6. Discussion. In summary, we show in this paper that the adjacency spectral embedding can be used to generate simple and intuitive test statistics for the semiparametric inference problem of testing whether two random dot product graphs on the same vertex set have the same or related generating latent positions.

The test statistic based on orthogonal Procrustes matching $\min_{\mathbf{W} \in \mathcal{O}(d)} \|\hat{\mathbf{X}}\mathbf{W} - \hat{\mathbf{Y}}\|_F$ is but one of many possible test statistics for testing the hypothesis $\mathbf{X} \perp \mathbf{Y}$. For example, the test statistic $\|\mathbf{A} - \mathbf{B}\|_F$ is intuitively appealing; it is a surrogate measure for the difference $\|\mathbf{X}\mathbf{X}^T - \mathbf{Y}\mathbf{Y}^T\|_F$. Furthermore, $\|\mathbf{A} - \mathbf{B}\|_F^2 = 2 \sum_{i < j} (\mathbf{A}_{ij} - \mathbf{B}_{ij})^2$ is a sum of independent Bernoulli random variables; hence it is easily analyzable and may possibly yield more powerful test. However, since $(\mathbf{A}_{ij} - \mathbf{B}_{ij})^2$ is

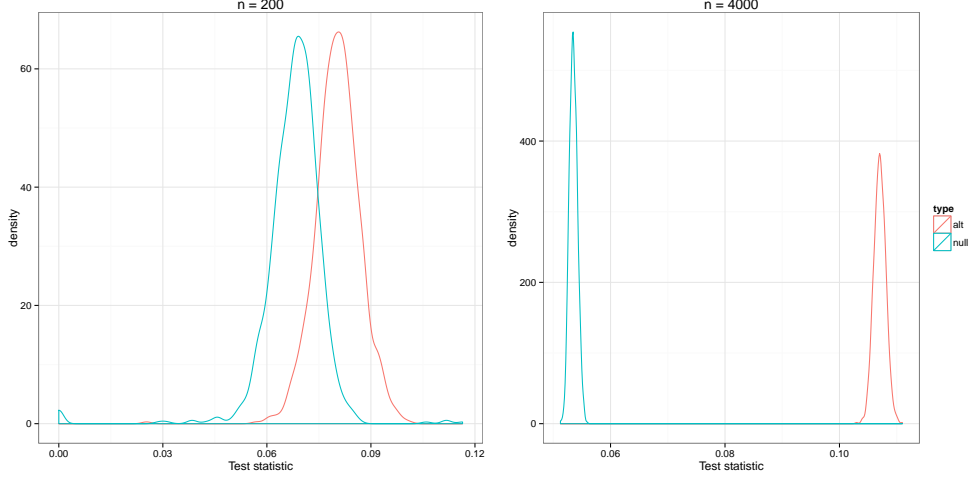


Fig 4: Density estimate for the test statistic when testing $H_0: \mathbf{X}_n \perp \mathbf{D}_n \mathbf{Y}_n$ for some diagonal matrix \mathbf{D}_n against the alternative $H_A: \mathbf{X}_n \not\perp \mathbf{D}_n \mathbf{Y}_n$ for all diagonal matrix \mathbf{D}_n .

a Bernoulli random variable with parameter $\mathbf{P}_{ij}(1-\mathbf{Q}_{ij}) + (1-\mathbf{P}_{ij})\mathbf{Q}_{ij}$, this forces that $(\mathbf{A}_{ij} - \mathbf{B}_{ij})^2 \sim \text{Bernoulli}(1/2)$ if $\mathbf{Q}_{ij} = 1/2$, regardless of the value of \mathbf{P}_{ij} . Therefore, $\|\mathbf{A} - \mathbf{B}\|_F^2 \sim \text{Binomial}(\binom{n}{2}, 1/2)$ whenever $\mathbf{Q} = 1/2\mathbf{J}$ where \mathbf{J} is the matrix of all ones. Thus, $\|\mathbf{A} - \mathbf{B}\|_F$ does not yield a consistent test; its distribution does not depend on \mathbf{P} whenever $\mathbf{B} \sim \text{Bernoulli}(1/2\mathbf{J})$.

Yet another simple test statistic is based on the spectral norm difference $\|\mathbf{A} - \mathbf{B}\|$; this is once again a surrogate measure for the difference $\|\mathbf{X}\mathbf{X}^T - \mathbf{Y}\mathbf{Y}^T\|$, and we surmise that such a test statistic may be more robust to model misspecification; e.g. when \mathbf{A} and \mathbf{B} are adjacency matrices of more general latent position random graphs. The concentration bound of [18], which we state in Eq. (2.1) in Proposition 2.3, can be used to construct a level- α test for the hypothesis $\mathbf{X} \perp \mathbf{Y}$. This may, however, lead to a test that is consistent for a narrower class of alternatives than that given in Theorem 3.1. Indeed, by Eq. (2.1), the test $\|\mathbf{A} - \mathbf{B}\|$ is consistent when the sequence $((\delta(\mathbf{P}_n) + \delta(\mathbf{Q}_n)) \log n)^{-1/2} \|\mathbf{P}_n - \mathbf{Q}_n\|$ diverges. In addition, we have

$$\begin{aligned} \|\mathbf{P}_n - \mathbf{Q}_n\| &= \|\mathbf{X}_n \mathbf{X}_n^T - \mathbf{Y}_n \mathbf{Y}_n^T\| \\ &= \|(\mathbf{X}_n \mathbf{W}_n - \mathbf{Y}_n)(\mathbf{X}_n \mathbf{W}_n + \mathbf{Y}_n)^T + (\mathbf{X}_n \mathbf{W}_n + \mathbf{Y}_n)(\mathbf{X}_n \mathbf{W}_n - \mathbf{Y}_n)^T\| \\ &\leq 2\|\mathbf{X}_n \mathbf{W}_n - \mathbf{Y}_n\| \|\mathbf{X}_n \mathbf{W}_n + \mathbf{Y}_n\| \leq 2\|\mathbf{X}_n \mathbf{W}_n - \mathbf{Y}_n\| \sqrt{\delta(\mathbf{P}_n) + \delta(\mathbf{Q}_n)}. \end{aligned}$$

for all orthogonal \mathbf{W}_n . Therefore, the fact that $(\delta(\mathbf{P}_n) + \delta(\mathbf{Q}_n))^{-1/2} \|\mathbf{P}_n - \mathbf{Q}_n\|$ diverges implies that $\min_{\mathbf{W}_n \in \mathcal{O}(d)} \|\mathbf{X}_n \mathbf{W}_n - \mathbf{Y}_n\|$ diverges, but not conversely, and hence $\|\mathbf{A} - \mathbf{B}\|$ may be consistent for a narrower class of alternatives. In addition, test statistics based directly on the adjacency matrices are also less flexible. For instance, it is not obvious to us that such test statistics can be easily adapted to test the hypothesis $\mathbf{X} \perp \mathbf{D}\mathbf{Y}$ for some diagonal matrix \mathbf{D} , or to conduct the nonparametric test of equality of the underlying distributions for the latent positions.

Finally, test statistics based on the spectral decomposition of the normalized Laplacian matrices can also be constructed. However, the resulting embedding is an estimate of some transformation of the latent positions rather than the latent positions themselves. More specifically, denote by $\tilde{\mathbf{X}}_n$ and $\tilde{\mathbf{Y}}_n$ the spectral decomposition obtained from the normalized Laplacian matrices associated with \mathbf{A}_n and \mathbf{B}_n , respectively. Then $\tilde{\mathbf{X}}_n$ is, up to some orthogonal transformation, “close” to $\mathcal{L}(\mathbf{X}_n)$

where $\mathcal{L}(\mathbf{X}_n)$ is a transformation of \mathbf{X}_n , i.e., the i -th row of $\mathcal{L}(\mathbf{X}_n)$ is given by $X_i/\langle X_i, \sum_{j \neq i} X_j \rangle$; similarly, $\tilde{\mathbf{Y}}_n$ is “close” to $\mathcal{L}(\mathbf{Y}_n)$. The construction of test statistics for testing the hypothesis in Section 2 for \mathbf{X}_n and \mathbf{Y}_n based on the estimates $\tilde{\mathbf{X}}_n$ and $\tilde{\mathbf{Y}}_n$ of $\mathcal{L}(\mathbf{X}_n)$ and $\mathcal{L}(\mathbf{Y}_n)$ is certainly possible; however, subtle technical issues regarding assumptions on the sequence of latent positions and speed of convergence of the estimates $\tilde{\mathbf{X}}_n$ and $\tilde{\mathbf{Y}}_n$ can arise. In summary, the formulation of the hypotheses and the accompanying test procedures in Section 2 are such that the test statistics are simple functions of the adjacency spectral embeddings of the graphs. Other formulations of comparable two-sample tests could, of course, lead to test statistics that are simple functions of the normalized Laplacian embeddings.

REFERENCES

- [1] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis* **11** 581–598.
- [2] ATHREYA, A., LYZINSKI, V., MARCHETTE, D. J., PRIEBE, C. E., SUSSMAN, D. L. and TANG, M. (2013). A limit theorem for scaled eigenvectors of random dot product graphs. Arxiv preprint. <http://arxiv.org/abs/1305.7388>.
- [3] BICKEL, P. and SARKAR, P. (2013). Role of normalization for spectral clustering in stochastic blockmodels. Arxiv preprint. <http://arxiv.org/abs/1310.1495>.
- [4] BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences of the United States of America* **106** 21068–73.
- [5] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A nonasymptotic theory of independence*. Oxford University Press.
- [6] CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral partitioning of graphs with general degrees and the extended planted partition model. In *Proceedings of the 25th conference on learning theory*.
- [7] CHOI, D. S., WOLFE, P. J. and AIROLDI, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika* **99** 273–284.
- [8] CONTE, D., FOGGIA, P., SANSONE, C. and VENTO, M. (2004). Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence* **18** 265–298.
- [9] DAVIS, C. and KAHAN, W. (1970). The rotation of eigenvectors by a perturbation. III. *Siam Journal on Numerical Analysis* **7** 1–46.
- [10] FORTUNATO, S. (2010). Community detection in graphs. *Physics Reports* **486** 75–174.
- [11] GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning* **2** 129–233.
- [12] HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97** 1090–1098.
- [13] HOLLAND, P. W., LASKEY, K. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5** 109–137.
- [14] HOOVER, D. N. (1979). Relations on probability spaces and arrays of random variables Technical Report, Institute for Advanced Study.
- [15] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E* **83** 016107, 10.
- [16] LYZINSKI, V., SUSSMAN, D. L., TANG, M., ATHREYA, A. and PRIEBE, C. E. (2013). Perfect Clustering for Stochastic Blockmodel Graphs via Adjacency Spectral Embedding. Arxiv preprint. <http://arxiv.org/abs/1310.0532>.
- [17] NEWMAN, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* **103** 8577–8582.
- [18] OLIVEIRA, R. I. (2010). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. Arxiv preprint <http://arxiv.org/abs/0911.0600>.
- [19] QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. *Advances in Neural Information Processing Systems*.
- [20] ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics* **39** 1878–1915.
- [21] SNIJDERS, T. A. B. and NOWICKI, K. (1997). Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure. *Journal of Classification* **14** 75–100.

- [22] SUSSMAN, D. L., TANG, M., FISHKIND, D. E. and PRIEBE, C. E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association* **107** 1119–1128.
- [23] SUSSMAN, D. L., TANG, M. and PRIEBE, C. E. (2014). Consistent latent position estimation and vertex classification for random dot product graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36** 48–57.
- [24] TANG, M., SUSSMAN, D. L. and PRIEBE, C. E. (2013). Universally consistent vertex classification for latent position graphs. *Annals of Statistics* **31** 1406–1430.
- [25] VOGELSTEIN, J. T., CONROY, J. M., PODRAZIK, L. J., KRATZER, S. G., HARLEY, E. T., FISHKIND, D. E., VOGELSTEIN, R. J. and PRIEBE, C. E. Fast approximate quadratic programming for large (brain) graph matching. Arxiv preprint. <http://arxiv.org/abs/1112.5507>.
- [26] WANG, H., TANG, M., PARK, Y. and PRIEBE, C. E. (2014). Locality statistics for anomaly detection in time series of graphs. *IEEE Transactions on Signal Processing* **62** 703–717.
- [27] YOUNG, S. and SCHEINERMAN, E. (2007). Random dot product graph models for social networks. In *Proceedings of the 5th international conference on algorithms and models for the web-graph* 138–149.
- [28] ZASLAVSKIY, M., BACH, F. and VERT, J. P. (2009). A path following algorithm for the graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31** 2227–2242.

DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS,
JOHNS HOPKINS UNIVERSITY,
3400 N. CHARLES ST, BALTIMORE, MD 21218, USA.
E-MAIL: mtang10@jhu.edu
dathrey1@jhu.edu
dsussma3@jhu.edu
vlyzins1@jhu.edu
cep@jhu.edu